The Structure Of An Atom Feed

Think of a weblog, or in fact any website with frequently updated content, like CNN.com. The site itself has a title ("CNN.com"), a subtitle ("Breaking News, U.S., World, Weather, Entertainment & Video News"), a last-updated date ("updated 12:43 p.m. EDT, Sat May 16, 2009"), and a list of articles posted at different times. Each article also has a title, a first-published date (and maybe also a last-updated date, if they published a correction or fixed a typo), and a unique URL.

The Atom syndication format is designed to capture all of this information in a standard format. My weblog and CNN.com are wildly different in design, scope, and audience, but they both have the same basic structure. CNN.com has a title; my blog has a title. CNN.com publishes articles; I publish articles.

At the top level is the *root element*, which every Atom feed shares: the `feed` element in the `http://www.w3.org/2005/Atom` namespace.

```
<feed xmlns='http://www.w3.org/2005/Atom'   !-- ① -->
      xml:lang='en'>                        <!-- ② -->
```

① `http://www.w3.org/2005/Atom` is the Atom namespace.

② Any element can contain an `xml:lang` attribute, which declares the language of the element and its children. In this case, the `xml:lang` attribute is declared once on the root element, which means the entire feed is in English.

An Atom feed contains several pieces of information about the feed itself. These are declared as children of the root-level feed element.

```
<feed xmlns='http://www.w3.org/2005/Atom' xml:lang='en'>
  <title>dive into mark</title>                              <!-- ① -->
  <subtitle>currently between addictions</subtitle>          <!-- ② -->
  <id>tag:diveintomark.org,2001-07-29:/</id>                 <!-- ③ -->
  <updated>2009-03-27T21:56:07Z</updated>                    <!-- ④ -->
```

```
<link rel='alternate' type='text/html' href='http://diveintomark.org/'/>  <!-- ⑤ -->
```

① The title of this feed is `dive into mark` .

② The subtitle of this feed is `currently between addictions` .

③ Every feed needs a globally unique identifier. See RFC 4151 for how to create one.

④ This feed was last updated on March 27, 2009, at 21:56 GMT. This is usually equivalent to the last-modified date of the most recent article.

⑤ Now things start to get interesting. This link element has no text content, but it has three attributes: `rel` , `type` , and `href` . The `rel` value tells you what kind of link this is; `rel='alternate'` means that this is a link to an alternate representation of this feed. The `type='text/html'` attribute means that this is a link to an HTML page. And the link target is given in the `href` attribute.

Now we know that this is a feed for a site named "dive into mark" which is available at http://diveintomark.org/ and was last updated on March 27, 2009.

> Although the order of elements can be relevant in some XML documents, it is not relevant in an Atom feed.

After the feed-level metadata is the list of the most recent articles. An article looks like this:

```
<entry>
  <author>                                                           <!-- ① -->
    <name>Mark</name>
    <uri>http://diveintomark.org/</uri>
  </author>
  <title>Dive into history, 2009 edition</title>                     <!-- ② -->
  <link rel='alternate' type='text/html'
    href='http://diveintomark.org/archives/2009/03/27/dive-into-history-2009-edition'/> <!--
  <id>tag:diveintomark.org,2009-03-27:/archives/20090327172042</id>  <!-- ④ -->
  <updated>2009-03-27T21:56:07Z</updated>                            <!-- ⑤ -->
  <published>2009-03-27T17:20:42Z</published>
  <category scheme='http://diveintomark.org' term='diveintopython'/> <!-- ⑥ -->
  <category scheme='http://diveintomark.org' term='docbook'/>
  <category scheme='http://diveintomark.org' term='html'/>
  <summary type='html'>Putting an entire chapter on one page sounds   <!-- ⑦ -->
    bloated, but consider this &amp;mdash; my longest chapter so far
    would be 75 printed pages, and it loads in under 5 seconds&amp;hellip;
    On dialup.</summary>
</entry>
```

① The `author` element tells who wrote this article: some guy named Mark, whom you can find loafing at http://diveintomark.org/. (This is the same as the alternate link in the feed metadata, but it doesn't have to be. Many weblogs have multiple authors, each with their own personal website.)

② The `title` element gives the title of the article, "Dive into history, 2009 edition".

③ As with the feed-level alternate link, this `link` element gives the address of the HTML version of this article.

④ Entries, like feeds, need a unique identifier.

⑤ Entries have two dates: a first-published date (`published`) and a last-modified date (`updated`).

⑥ Entries can have an arbitrary number of categories. This article is filed under `diveintopython`, `docbook`, and `html`.

⑦ The `summary` element gives a brief summary of the article. (There is also a `content` element, not shown here, if you want to include the complete article text in your feed.) This `summary` element has the Atom-specific `type='html'` attribute, which specifies that this summary is a snippet of html, not plain text. This is important, since it has html-specific entities in it (`—` and `…`) which should be rendered as "—" and "…" rather than displayed directly.

⑧ Finally, the end tag for the `entry` element, signaling the end of the metadata for this article.