The main building blocks of markup languages are elements expressed in tags. For example, HTML markup has at its base the <HTML>…</HTML> element. Of course, any number of other elements can be contained within that base.

Valid XML documents — and HTML documents when seen through the lens of a DTD — are made up by the following building blocks: Elements, tags, attributes, entities, PCDATA and CDATA

Elements

Elements are the main building blocks of both XML and HTML documents. Examples of HTML elements include "body" and "table". Examples of XML elements could be "memo" and "message". Elements can contain text, other elements or remain empty. Examples of empty HTML elements are "hr", "br" and "img".

Tags

Tags are used to mark up elements. A starting tag like <element_name> marks up the beginning of an element, and an ending tag like </element_name> marks up the end of an element. Examples:

Code	Explanation
`<body>body text in between</body>.`	Body element marked up with body tags
`<message> message in between</message>`	Message element marked up with message tags

Attributes

Attributes provide extra information about elements. They are always placed inside the starting tag of an element. Attributes always come in name/value pairs. The following "img" element has additional information about a source file:

```
<img src="mt_mazama.gif" />
```

The name of the element is "img". The name of the attribute is "src". The value of the attribute is "mt_mazama.gif". Since the element itself is empty it is closed by a "/".

Entities

An entity is a reference to other data that often acts as an abbreviation or a shortcut. By declaring entities, you can avoid entering the same information in a document or DTD repetitively. Entities are variables used to define common text. The HTML code for a “non-breaking space” is the entity “ .” XML uses the same pre-defined entities. Entity references are references to entities.

Entities are expanded when a document is parsed by an XML parser. Think of them as sponges that are hard and small when dry but expand to full bloom when water is added. Here is an example:

Code	Explanation
`<!ENTITY hccs “Houston Community College System”>`	The abbreviation “hccs” can be used in XML code while allowing the parser to expand it into “Houston Community College System” as shown below.
`<description>&hccs; is a comprehensive community college offering a wide variety of academic and workforce programs ...</description>`	Houston Community College System is a comprehensive community college offering a wide variety of academic and workforce programs ...

The following are pre-defined internal general entities aimed at allowing the display of characters that may not be available because of encoding restrictions of character usage in programming. Although the list of HTML entities is long, the following five are built into XML to aid the developer in avoiding markup problems.

Entity Reference			Character
"		"	#
&	*or*	&	&
<	*or*	<	<
>	*or*	>	>
		'	‘

PCDATA

PCDATA signifies “parsed character data.” Character data is text found between the start tag and the end tag of an XML element. PCDATA is text that will be interpreted by a parser. Tags inside the text will be treated as markup and entities will be expanded.

CDATA

CDATA means “character data.” It is text that will not be interpreted by a parser. Tags inside the text will not be treated as markup and entities will not be expanded.