Синтаксис тегов

Обычно узел XML-документа задается парой тегов. Тег — это некоторый идентификатор, показывающий тип содержимого, ограниченный символом “<” и “>”. Тег бывает начальным и конечным. Идентификатору в конечном теге обязательно должен предшествовать символ “/”:

<name>Jonh</name>

Дерево XML-документа всегда завершается пустыми узлами. Это такие узлы, у которых нет дочерних узлов. Соответственно, их теги содержат только текст, но не другие теги. Если же в пустом узле нет и текста, то использовать для его задания пару тегов не нужно (но допустимо). Его можно инициализировать одинарными тегами, начинающимися с “<” и заканчивающимися “/>”. Например:

<person>

<name/>

<age/>

</person>

Одинарные теги обычно применяются в сочетании с атрибутами (см. ниже). Для задания имен узлов можно использовать любые буквы и цифры. Также в них могут присутствовать три символа пунктуации: “_” (подчеркивание), “-” (дефис), “.” (точка) (однако в начале имени из этих знаков может стоять только “_”). Другие знаки, а также пробельные символы входить в имя не могут. Ограничений на длину имени не существует. Имена узлов чувствительны к регистру.

Атрибуты

Узлы XML-дерева могут иметь собственные свойства, которые принято называть атрибутами. Задаются атрибуты при помощи пар имя-значение, помещенных в начальный тег. Значение атрибута окружается одинарными или двойными кавычками. Правила задания имен для атрибутов те же, что и для узлов. Например, следующий текст задает корневой узел person, имеющий атрибуты name и age:

```
<person name=’Jonh’ age=’23’></person>
```

Используя атрибуты, зачастую можно вообще обойтись без вложения в тег текста и создания части дочерних узлов. Однако полностью заменить атрибутами вложенное содержимое узла невозможно. Во-первых, каждый атрибут должен иметь уникальное имя. Это не позволяет создавать при помощи них перечисления, которые легко реализуются посредством группы одинаковых тегов. Во-вторых, значение атрибута не разбирается анализатором, поэтому его текст не может иметь разметки. Текст же, вложенный в тег, может содержать другие теги и прочие служебные структуры. В общем, причина ограниченности атрибутов та же, которая делает невозможным описание сложного документа посредством строки с парами имя-значение. Обычно считается, что в атрибутах указываются метаданные элемента, в то время как сама информация, для хранения которой был создан XML-документ, должна помещаться внутрь тега. Впрочем, подобное деление довольно условно. На практике же атрибуты обычно используются в пустых узлах, завершающих дерево документа.