Лекция 6

Язык разметки XML. Структура документа

XML ничего не делает сам

```xml
<?xml version="1.0" encoding="UTF-8"?>
<note>
   <to>Виктория</to>
   <from>Кирилл</from>
   <heading>Напоминание</heading>
   <body>Не забудь сделать уроки!</body>
</note>
```

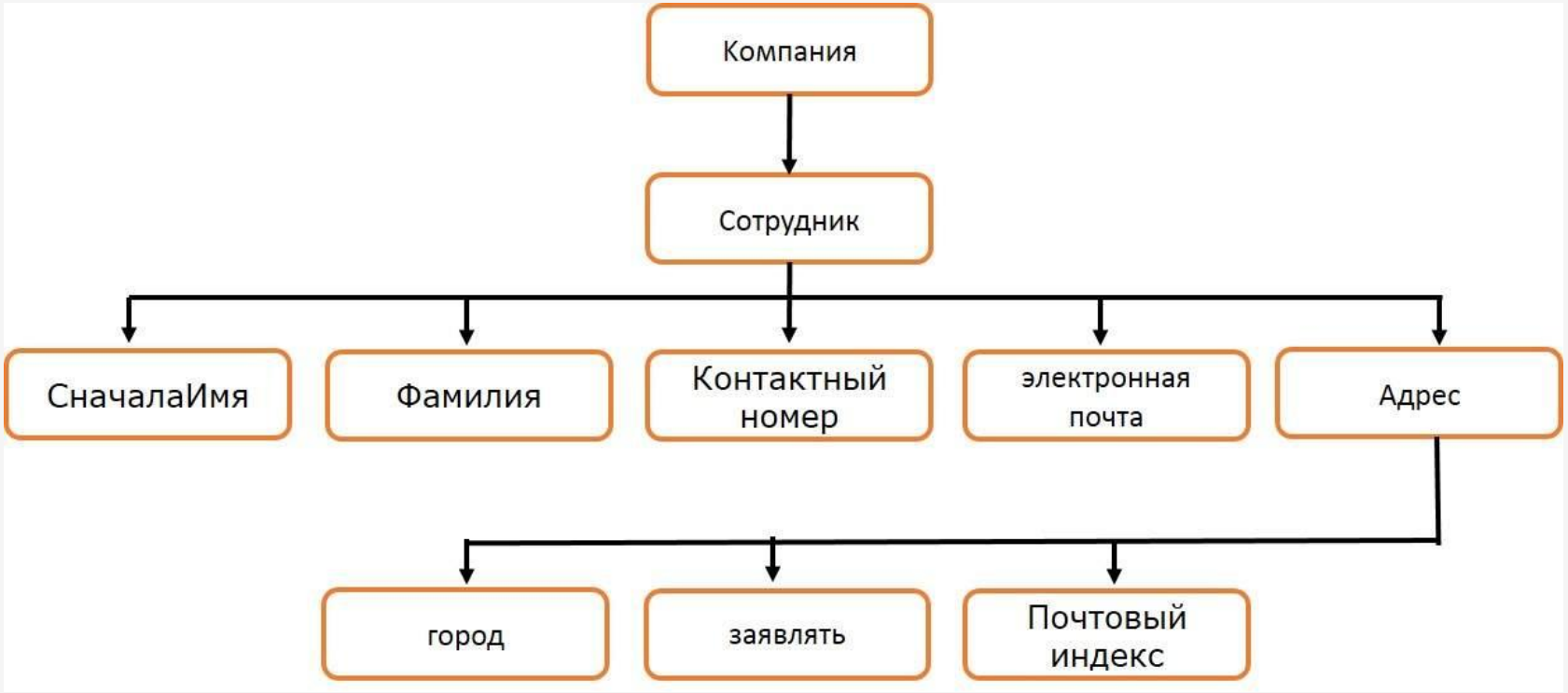
```xml
<?xml version="1.0" encoding="UTF-8"?>
<note>…
</note>
```

Вы сами изобретаете теги:

- <note> - заметка
- <to> - кому
- <from> - от кого
- И т.д.

Тег - идентификатор для категоризации, описания, поиска данных и задания внутренней структуры. В языке XML заключается в треугольные скобки - **<тег>**

XML имеет древовидную структуру



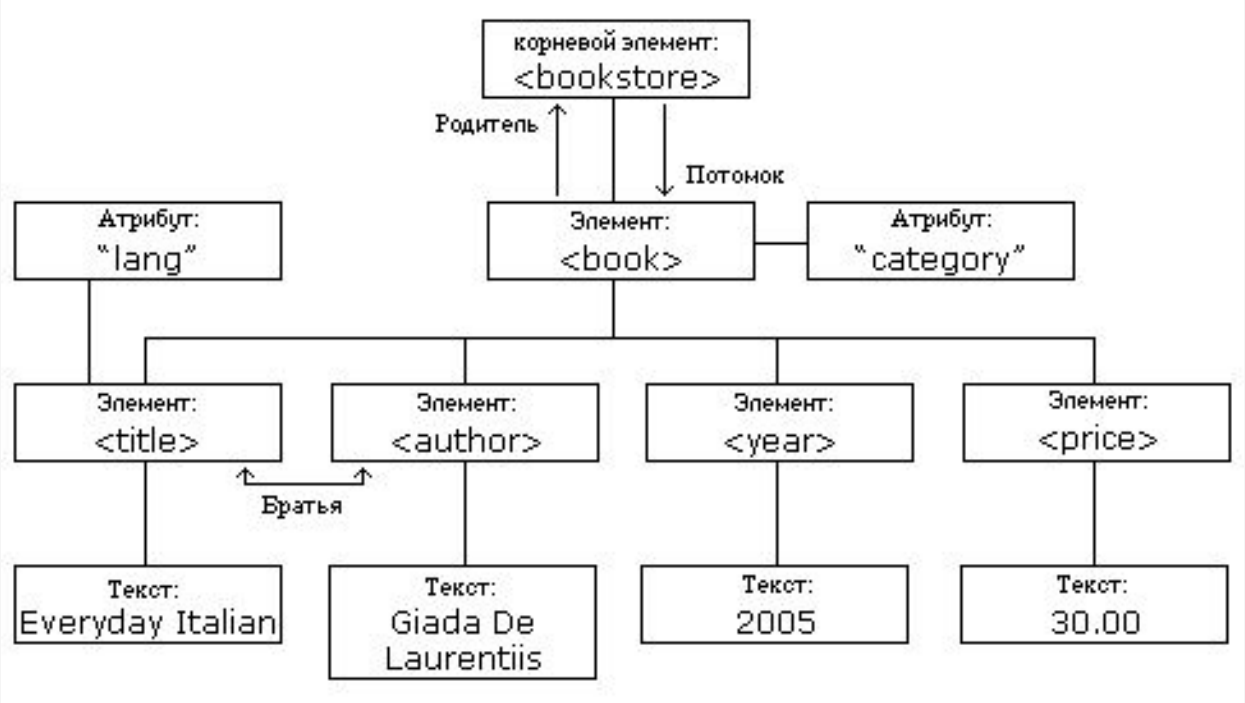
Разберем детальнее

```xml
<?xml version="1.0" encoding="UTF-8"?>
<note>
    <to>Виктория</to>
    <from>Кирилл</from>
    <heading>Напоминание</heading>
    <body>Не забудь сделать уроки!</body>
</note>
```

XML всегда придерживается древовидной структуры

```
<?xml version="1.0" encoding="UTF-8"?>
<корневой>
    <потомок>
        <подпотомок>.....</подпотомок>
    </потомок>
</корневой>
```

Отношение в XML – родитель, дочерний элемент, брат/сестра



Реализация XML

```xml
<?xml version="1.0" encoding="UTF-8"?>
<bookstore>
    <book category="COOKING">
        <title lang="en">Everyday Italian</title>
        <author>Giada De Laurentiis</author>
        <year>2005</year>
        <price>30.00</price>
    </book>
    <book category="CHILDREN">
        <title lang="en">Harry Potter</title>
        <author>J K. Rowling</author>
        <year>2005</year>
        <price>29.99</price>
    </book>
    <book category="WEB">
        <title lang="en">Learning XML</title>
        <author>Erik T. Ray</author>
        <year>2003</year>
        <price>39.95</price>
    </book>
</bookstore>
```

Все XML элементы должны иметь закрывающий тег

```xml
<?xml version="1.0" encoding="UTF-8"?>

<!-- неправильно -->
<info>Привет всем!

<!-- правильно -->
<info>Привет всем!</info>
```

Теги XML регистрозависимы

Тег **<Letter>** не то же самое, что тег **<letter>**.

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- неправильно -->
<Message>Это неправильно</message>

<!-- правильно -->
<message>Это правильно</message>
```

Соблюдайте корректную вложенность

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- неправильно -->
<b><i>Это жирный и курсивный текст</b></i>

<!-- правильно -->
<b><i>Это жирный и курсивный текст</i></b>
```

У XML документов всегда один общий родительский элемент – корневой элемент

```
<?xml version="1.0" encoding="UTF-8"?>
<корневой>
    <потомок>
        <подпотомок>.....</подпотомок>
    </потомок>
</корневой>
```

XML пролог – необязателен, но рекомендован и всегда на первом месте

```xml
<?xml version="1.0" encoding="UTF-8"?>
<note>
   <to>Tove</to>
   <from>Jani</from>
   <heading>Напоминание</heading>
   <body>Не забудь сделать уроки!</body>
</note>
```

Правильно

```xml
<note>
   <to>Tove</to>
   <from>Jani</from>
   <heading>Напоминание</heading>
   <?xml version="1.0" encoding="UTF-8"?>
   <body>Не забудь сделать уроки!</body>
</note>
```

Неправильно

Атрибуты XML только в кавычках

```xml
<?xml version="1.0" encoding="UTF-8"?>

<!-- неправильно -->
<note date=12/11/2007>
    <to>Tove</to>
    <from>Jani</from>
</note>

<!-- правильно -->
<note date="12/11/2007">
    <to>Tove</to>
    <from>Jani</from>
</note
```

```xml
<?xml version="1.0" encoding="UTF-8"?>

<!-- неправильно -->
<note date=12/11/2007>…
</note>

<!-- правильно -->
<note date="12/11/2007">…
</note
```

Сущность – символ с особым значением

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- неправильно -->
<message>если жалование < 1000</message>


<!-- правильно -->
<message>если жалование &lt; 1000</message>
```

Sublime Text подчеркнул ошибку для нашего удобства

Все сущности

Сущность	Символ	Значение
<	<	меньше, чем
>	>	больше, чем
&	&	амперсанд
'	'	апостроф
"	"	кавычки

Только символы "<" и "&" строго запрещены в XML. Символ ">" допустим, но лучше его всегда заменять на сущность

Комментарии в XML – текст, который не превращается в разметку

```
<!-- Это комментарий -->

<!-- Это -- ошибка -->

<!-- Это - - комментарий -->
```

Элементы XML:

- Текст
- Атрибуты

```
<bookstore>
    <book category="CHILDREN">
        <title>Harry Potter</title>
        <author>J K. Rowling</author>
        <year>2005</year>
        <price>29.99</price>
    </book>
    <book category="WEB">
        <title>Learning XML</title>
        <author>Erik T. Ray</author>
        <year>2003</year>
        <price>39.95</price>
    </book>
</bookstore>
```

Пустые XML элементы

```xml
<?xml version="1.0" encoding="UTF-8"?>

<!-- обычный синтаксис -->
<book></book>

<!-- самозакрывающийся тег -->
<book />
```

Правила написания имен XML

- Имена могут содержать буквы, числа и другие символы
- Имена не могут начинаться с цифры или символа пунктуации
- Имена не могут начинаться с сочетания "xml" (или XML, или Xml и т.п.)
- Имена не могут содержать пробельные символы

Хорошая практика составления имен

- Старайтесь придумать описательные имена: **<first_name>, <last_name>**.
- Имена следует составлять короткие и простые, вроде: **<book_title>**; а не: **<the_title_of_the_book>**.
- Избегайте символ "-". Если вы напишите нечто вроде "first-name", то некоторые приложения могут решить, что вы вычитаете имя "name" из имени "first".
- Избегайте символ ".". Если вы напишите нечто вроде "first.name", то некоторые приложения могут решить, что "name" это свойство объекта "first".
- Избегайте символ ":". Двоеточие зарезервировано для механизма пространства имен.

Стили написания имен

Стиль	Пример	Описание
Нижний регистр	<firstname>	Все буквы в нижнем регистре
Верхний регистр	<FIRSTNAME>	Все буквы в верхнем регистре
С символом подчеркивания	<first_name>	Слова разделяются символом подчеркивания
В стиле Pascal	<FirstName>	Первые буквы всех слов в верхнем регистре
"Верблюжий горб« -camelcase	<firstName>	Первые буквы всех слов за исключением первого в верхнем регистре

Расширяемость – главная особенность XML

```xml
<?xml version="1.0" encoding="UTF-8"?>

<note>
    <to>Tove</to>
    <from>Jani</from>
    <body>Не забудь про уборку!</body>
</note>
```

Сообщение:
Кому: Tove
От кого: Jani
Текст: Не забудь про уборку!

Добавление новых полей не повредит старую структуру

```xml
<?xml version="1.0" encoding="UTF-8"?>

<note>
   <date>2008-01-10</date>
   <to>Tove</to>
   <from>Jani</from>
   <heading>Напоминание</heading>
   <body>Не забудь про уборку!</body>
</note>
```

Сообщение:
Кому: Tove
От кого: Jani
Текст: Не забудь про уборку!

Атрибуты – дополняют поля информацией

```
<?xml version="1.0" encoding="UTF-8"?>

<imgage src='computer.gif'>Картинка</img>
<text size='12px'>Привет всем!</text>
<file type="gif">computer.gif</file>
```

Атрибуты в кавычках

- “” - двойные
- ‘’ - одинарные

```
<?xml version="1.0" encoding="UTF-8"?>

<person sex="female">
<person sex='female'>
<gangster name='George "Shotgun" Ziegler'>
<gangster name="George &quot;Shotgun&quot; Ziegler">
```

Атрибуты или элементы?

```xml
<?xml version="1.0" encoding="UTF-8"?>

<person sex="female">
    <firstname>Anna</firstname>
    <lastname>Smith</lastname>
</person>
```

```xml
<?xml version="1.0" encoding="UTF-8"?>

<person>
    <sex>female</sex>
    <firstname>Anna</firstname>
    <lastname>Smith</lastname>
</person>
```

Данные примеры идентичны

```xml
<?xml version="1.0" encoding="UTF-8"?>

<note date="10/01/2008">
   <to>Tove</to>
   <from>Jani</from>
   <heading>Напоминание</heading>
   <body>Не забудь обо мне в эти выходные!</body>
</note>
```

```xml
<?xml version="1.0" encoding="UTF-8"?>

<note>
   <date>10/01/2008</date>
   <to>Tove</to>
   <from>Jani</from>
   <heading>Напоминание</heading>
   <body>Не забудь обо мне в эти выходные!</body>
</note>
```

Правильный вариант!

```xml
<?xml version="1.0" encoding="UTF-8"?>

<note>
    <date>
        <day>10</day>
        <month>01</month>
        <year>2008</year>
    </date>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Напоминание</heading>
    <body>Не забудь обо мне в эти выходные!</body>
</note>
```

Символьная кодировка определяет уникальный бинарный код для различных символов, используемых в документе.

Юникод — это промышленный стандарт для символьной кодировки текстового документа. Он определяет (почти) все возможные международные символы по именам и числам.

- UTF-8 – 8 бит для общепринятых символов (1 байт)
- UTF-16 – 16 бит для общепринятых символов (2 байта)

UTF-8 – это стандарт для WEB разработки и XML документов

```
<?xml version="1.0" encoding="UTF-8"?>
```

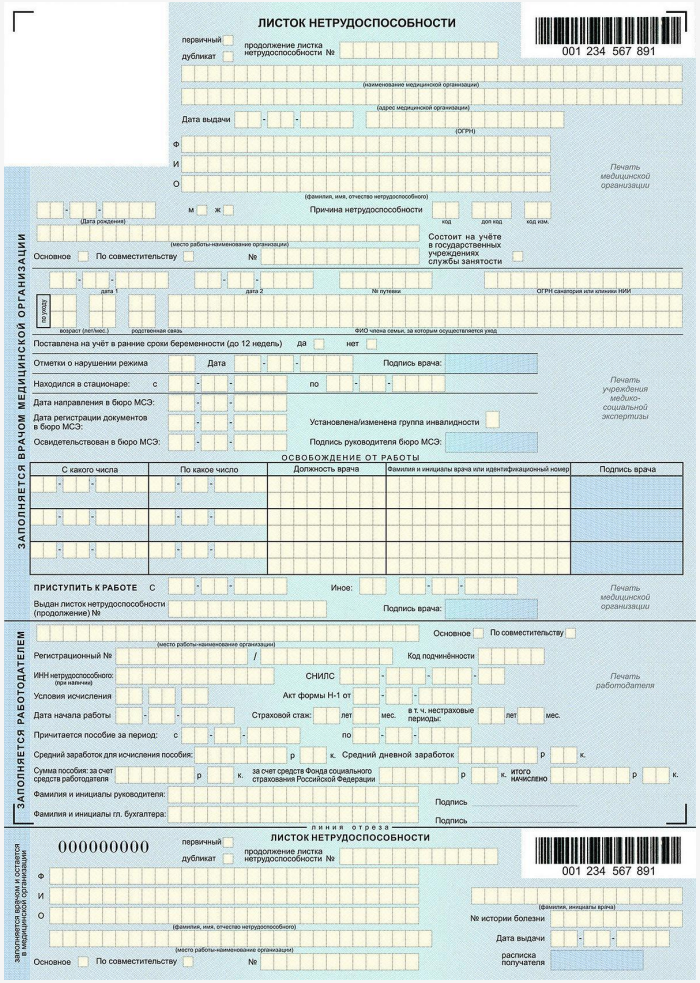
0	[NULL]	10	[DATA LINK ESCAPE]	20	[SPACE]	30	0	40	@	50	P	60	`	70	p
1	[START OF HEADING]	11	[DEVICE CONTROL 1]	21	!	31	1	41	A	51	Q	61	a	71	q
2	[START OF TEXT]	12	[DEVICE CONTROL 2]	22	"	32	2	42	B	52	R	62	b	72	r
3	[END OF TEXT]	13	[DEVICE CONTROL 3]	23	#	33	3	43	C	53	S	63	c	73	s
4	[END OF TRANSMISSION]	14	[DEVICE CONTROL 4]	24	$	34	4	44	D	54	T	64	d	74	t
5	[ENQUIRY]	15	[NEGATIVE ACKNOWLEDGE]	25	%	35	5	45	E	55	U	65	e	75	u
6	[ACKNOWLEDGE]	16	[SYNCHRONOUS IDLE]	26	&	36	6	46	F	56	V	66	f	76	v
7	[BELL]	17	[END OR TRANS. BLOCK]	27	'	37	7	47	G	57	W	67	g	77	w
8	[BACKSPACE]	18	[CANCEL]	28	(	38	8	48	H	58	X	68	h	78	x
9	[HORIZONTAL TAB]	19	[END OF MEDIUM]	29	)	39	9	49	I	59	Y	69	i	79	y
A	[LINE FEED]	1A	[SUBSTITUTE]	2A	*	3A	:	4A	J	5A	Z	6A	j	7A	z
B	[VERTICAL TAB]	1B	[ESCAPE]	2B	+	3B	;	4B	K	5B	[	6B	k	7B	{
C	[FORM FEED]	1C	[FILE SEPARATOR]	2C	,	3C	>	4C	L	5C	\	6C	l	7C	|
D	[CARRIAGE RETURN]	1D	[GROUP SEPARATOR]	2D	-	3D	=	4D	M	5D	]	6D	m	7D	}
E	[SHIFT OUT]	1E	[RECORD SEPARATOR]	2E	.	3E	>	4E	N	5E	^	6E	n	7E	~
F	[SHIFT IN]	1F	[UNIT SEPARATOR]	2F	/	3F	?	4F	O	5F	_	6F	o	7F	[DEL]

Просмотр файлов XML поддерживает любой браузер

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<note>
  ▼<date>
     <day>10</day>
     <month>01</month>
     <year>2008</year>
   </date>
   <to>Tove</to>
   <from>Jani</from>
   <heading>Напоминание</heading>
   <body>Не забудь обо мне в эти выходные!</body>
 </note>
```

Больничный в XML



Спасибо за внимание!