**Лабораторная работа № 11 Создание и валидация ХМL-документов.**

**Цель работы:** изучить правила создания XML-документов, получить навыки создания валидных документов и их проверка.

**Теоретические сведения для выполнения работы**

**Основы XML**

XML (eXtensible Markup Language) — расширяемый язык разметки, созданный для хранения, транспортировки и обмена данными. Он включает в себя элементы SGML и предназначен для определения HTML-подобных языков. В нем отсутствуют элементы SGML, не применимые к языкам типа HTML, а другие элементы упрощены, чтобы облегчить их понимание и использование. В языке XML нет предопределенных тегов, автор определяет свои языковые теги и свою структуру документа. Формально XML представляет собой набор правил для создания собственных языков разметки, а также чтения и написания документов на языке разметки.

Пример XML-документа

|  |
| --- |
| <?xml version="1.0" encoding="UTF-8"?>  <note>  <to>Анна</to>  <from>Дмитрий</from>  <heading>Напоминание</heading>  <body>Не забудь обо мне в эти выходные!</body> </note> |

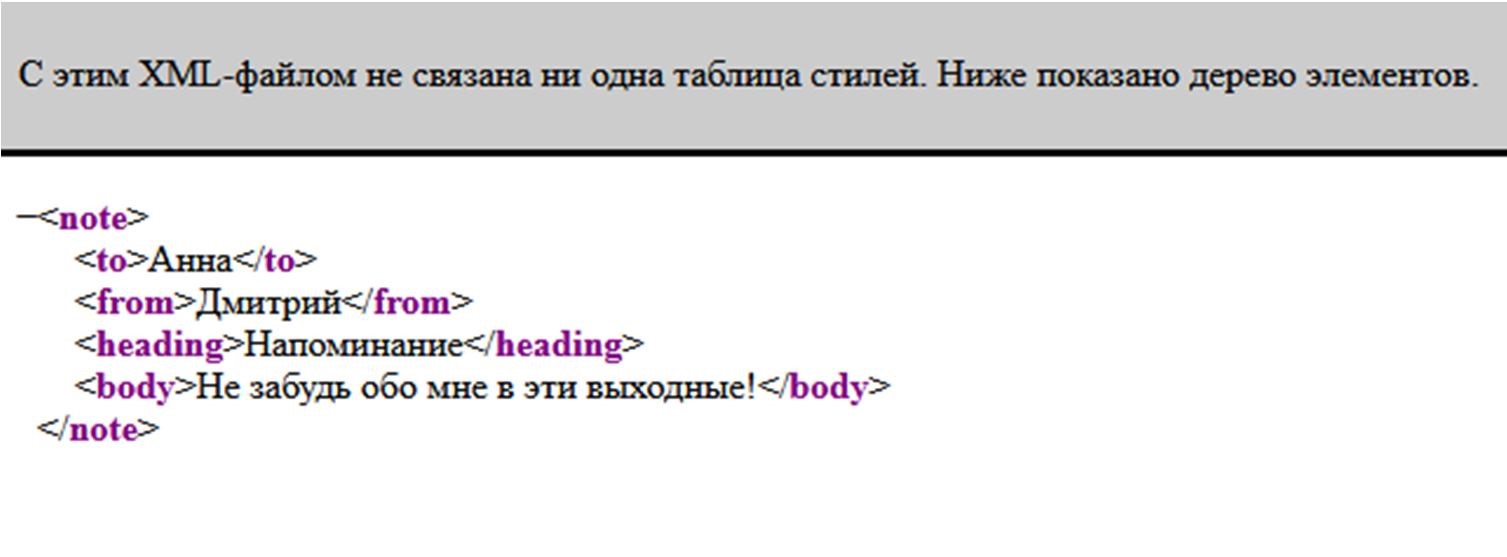
Согласно примеру XML-документ состоит из содержимого, заключенного между тегов. Элемент состоит из своего начального и конечного тегов и всего, что находится между ними. Первые строки называются прологом и объявляют процессору XML, что данный документ размечен в XML. Пролог включает объявление (XML-декларация), которое начинается с разделителя, состоящего из пяти символов **<?xml**, за которым следует некоторое количество определений свойств, каждое из которых состоит из имени свойства и его значения, заключенного в кавычки. Объявление завершается закрывающим разделителем из двух символов **?>**. Варианты объявлений могут быть разные в зависимости от предпочтений пользователя:

<?xml version="1.0"?>

<?xml version='1.0' encoding='UTF-8' standalone='yes'?>

<?xml version = '1.0' encoding= 'UTF-8' standalone ="no"?>

Элемент документа <note> является корневым. Следующие 4 строки описывают дочерние элементы корневого элемента: <to>Анна</to>, <from>Дмитрий</from>, <heading>Напоминание</heading>, <body>Не забудь обо мне в эти выходные!</body>. Последняя строка определяет конец корневого элемента: </note>. При открытии его в браузерах, например Mozilla или Opera, будет отображаться следующим образом:



Таким образом, необходимо подключение таблицы стилей для создания конечного продукта после объявления XML-документа.

<?xml-stylesheet type="text/css" href="style.css"?>

Пример XML-документа представлен в таблице 11.1

Таблица 11.1

|  |  |
| --- | --- |
| **XML-документ** | **CSS файл** |
| <?xmlversion="1.0" encoding="UTF-  8"?>  <!--FileName: 1.xml>  <?xml-stylesheet type="text/css"  href="1.css"?>  <FACULTY>  <SPECIALIZATION>  <NAME>  <TIME>  <EXAM>  <PAGES>  <PASSING>  </SPECIALIZATION>  <SPECIALIZATION>  <NAME>  <TIME>  <EXAM>  <PAGES>  <PASSING>  </SPECIALIZATION>  </FACULTY> | SPECIALIZATION  { display:block;  text-align:center; margin:10px 20px 10px 20px;  margin-top:12pt;  font-size:15pt}  NAME  { font-style:italic;  text-align:center}  TIME  { display:block; text-align:center;  font-weight:bold;  EXAM  { display:block; text-align:center;  }  PAGES  { display:block;  text-align:center;  PASSING  { display:block; text-align:center;  }  FACULTY  { display:block; margin:50px 100px; border-style:solid  border-color:#8C0088} |

Также следует различать синтаксически верный документ и валидные документы. К синтаксически верным (корректным) относятся XMLдокументы, которые соответствуют правилам синтаксиса XML:

1. Все XML элементы должны иметь закрывающий тег.
2. Теги XML являются регистрозависимыми.
3. Перед закрывающей угловой скобкой в пустых элементах XML требуется ставить косую черту.
4. Значения должны быть заключены в одинарные или двойные кавычки.
5. Все элементы обязаны соблюдать корректную вложенность.
6. XML документ должен содержать один корневой элемент, который будет родительским для всех других элементов.
7. Учитываются все символы форматирования (т.е. пробелы, переводы строк, табуляции не игнорируются, как в HTML).

В общем виде ХМL-документ после пролога можно представить в следующем виде:

<корневой>

<потомок>

<подпотомок>.....</подпотомок>

</потомок>

</корневой>

Имена элементов могут начинаться только с букв и символов подчеркивания и могут содержать только буквы, цифры, дефисы, точки и символы подчеркивания и не могут начинаться с сочетания «xml». В качестве имен можно использовать любые слова, зарезервированных слов нет.

**Валидные XML-документы**

Валидный XML-документ должен быть синтаксически верным и соответствовать одному из типов определения документов. Правила, определяющие допустимые элементы и атрибуты для XML документа, называются определениями документа или схемами документа.

С XML можно использовать различные типы определений документа:

* оригинальное определение типа документа (Document Type Definition, DTD).
* XML схема — тип определений, основанный на XML схема.

DTD (Document Type Definition, определение типа документа) — это язык описания структуры XML-документа, который используется для проверки граматики XML-документа и его соответствия определенному типу. Цель DTD состоит в том, чтобы определить структуру XML документа. Это делается путем определения списка допустимых элементов. В прологе декларация объявления элементов может быть внутренняя или внешняя

Таблица 11.2

|  |  |
| --- | --- |
| **Внутренняя декларация DTD** | **Внешняя декларация DTD** |
| <?xml version="1.0" encoding="UTF-8" ?>  <!DOCTYPE note [  <!ELEMENT note (to, from, Sbj, msg)>  <!ELEMENT to (#PCDATA)>  <!ELEMENT from (#PCDATA)>  <!ELEMENT Sbj (#PCDATA)>  <!ELEMENT msg (#PCDATA)>  ]> | <?xml version="1.0" encoding="UTF-8" ?>  <!DOCTYPE note SYSTEM "note.dtd">  <note>  <to>Sunny</to>  <from>Oliver</from>  <Sbj>Hello</Sbj>  <msg>This is a good day!</msg> </note> |
| <note>  <to>Sunny</to>  <from>Oliver</from>  <Sbj>Hello</Sbj>  <msg>This is a good day!</msg> </note> | **файл note.dtd содержит:**  <!ELEMENT note (to,from,heading,body)>  <!ELEMENT to (#PCDATA)>  <!ELEMENT from (#PCDATA)>  <!ELEMENT heading (#PCDATA)>  <!ELEMENT body (#PCDATA)> |

DTD в приведенном примере интерпретируется следующим образом:

1. !DOCTYPE note определяет, что корневым элементом документа является note
2. !ELEMENT note определяет, что элемент note содержит четыре

элемента: to, from, heading, body

1. !ELEMENT to определяет, что элемент to должен быть типа "#PCDATA"
2. !ELEMENT from определяет, что элемент from должен быть типа "#PCDATA"
3. !ELEMENT heading определяет, что элемент heading должен быть типа "#PCDATA"
4. !ELEMENT body определяет, что элемент body должен быть типа

"#PCDATA"

#PCDATA означает разбираемые текстовые данные. Кроме инструкции ELEMENT существуют следующие:

— ATTLIST для перечисления и объявления атрибутов, которые могут принадлежать элементу.

— ENTITY для определения сущностей в DTD с целью их использования как в связанном с DTD XML-документе, так и собственно в DTD.

|  |  |  |
| --- | --- | --- |
| <?xml version="1.0"?>  <!DOCTYPE note [  <!ENTITY name “Hello, world!”>  ]>  <element>&name;</element> | | |
|  | — NOTATION используется для определения нотаций. |  |

Сущность является заместителем содержания, которую можно однажды объявить и многократно использовать почти в любом месте документа. Встроенные сущности приведены в таблице 11.3

Таблица 11.3

|  |  |  |
| --- | --- | --- |
| **Сущность** | **Ссылка на сущность** | **Значение** |
| lt | &lt; | < (меньше чем) |
| gt | &gt; | > (больше чем) |
| amp | &amp; | & (амперсанд) |
| apos | &apos; | ' (апостроф или одиночная кавычка) |
| quot | &quot; | " (двойная кавычка) |

К основным недостаткам DTD схем относят синтаксические отличия от языка XML, а также отсутствие возможностей работы с типами данных. Для решения этих недостатков предлагается использовать XML Schema, которая описывают структуру XML документа. XML документ, прошедший проверку по XML схеме, является "синтаксически верным" и "валидным".

Кроме этого Schema имеет ряд преимуществ перед DTD

— XML схема пишется на XML;

— XML схема легко расширяется;

— XML схема поддерживает типы данных;

— XML схема поддерживает пространства имен;

Пример Schema представлен в таблице 11.4

Таблица 11.4

|  |  |  |
| --- | --- | --- |
| **XML-документ** | **XML Schema** | |
| <?xml version="1.0" encoding="UTF8"?> <note  xmlns:xsi="http://www.w3.org/2001/X MLSchema-instance"  xsi:noNamespaceSchemaLocation="not e.xsd">  <to>Sunny</to>  <from>Oliver</from>  <Sbj>Hello</Sbj>  <msg>This is a good day!</msg>  </note> | <?xml version="1.0" encoding="UTF  <xs:schema  xmlns:xs="http://www.w3.org/2001/XMLSchema">  <xs:element name="note"  <xs:complexType>  <xs:sequence>  <xs:element type="xs:string"/>  <xs:element type="xs:string"/>  <xs:element type="xs:string"/>  <xs:element type="xs:string"/>  </xs:sequence>  </xs:complexType>  </xs:element>  </xs:schema> | -8" ?>    >  name="to" name="from" name="heading" name="body" |

Приведенная выше схема интерпретируется следующим образом:

1. <xs:element name="note"> определяет элемент "note";
2. <xs:complexType> у элемента "note" комплексный тип;
3. <xs:sequence> последовательность элементов элемента ;
4. <xs:element name="to" type="xs:string"> у элемента "to" строковый тип (текст);
5. <xs:element name="from" type="xs:string"> у элемента "from" строковый тип;
6. <xs:element name="heading" type="xs:string"> у элемента "heading" строковый тип;
7. <xs:element name="body" type="xs:string"> у элемента "body" строковый тип.

Как видно из примера, каждая XML схема состоит с корневого элемента «schema» и обязательного пространства имен

«http://www.w3.org/2001/XMLSchema», а далее идет описание схемы.

**Задания к лабораторной работе № 11**

**Задание 1** Создать корректный XML-документ, который будет содержать информацию на произвольную тему. При выполнении задания используйте css.

**Задание 2** Преобразуйте XML-документ из задания 1 в XML DTD, предварительно сделав его копию. Должно быть внутреннее объявление DTD, в котором должны быть использованы сущности, которые затем применяются в XML-документе.

**Задание 3** Преобразуйте XML-документ из задания 1 в XML Schema, предварительно сделав его копию.

**Задания 4** Провести проверку валидности XML-документов, подтвердив скриншотами результатов.

**Примечание:** проверка валидности XML-документа

https://wtools.io/ru/validate-xml-online

пример XML Schema https://msiter.ru/tutorials/uchebnik-po-xml-

shemam/primer-xml-shemy для подготовки к защите воспользоваться из папки Литература 5 и 6 источником

**Контрольные вопросы**

1. Что такое XML?

XML (eXtensible Markup Language) — расширяемый язык разметки, созданный для хранения, транспортировки и обмена данными.

1. Чем является первая строка в коде XML-документа?

Первые строки называются прологом и объявляют процессору XML, что данный документ размечен в XML.

1. Назовите правила синтаксиса XML-документа?

XMLдокументы, которые соответствуют правилам синтаксиса XML:

1. Все XML элементы должны иметь закрывающий тег.
2. Теги XML являются регистрозависимыми.
3. Перед закрывающей угловой скобкой в пустых элементах XML требуется ставить косую черту.
4. Значения должны быть заключены в одинарные или двойные кавычки.
5. Все элементы обязаны соблюдать корректную вложенность.
6. XML документ должен содержать один корневой элемент, который будет родительским для всех других элементов.
7. Учитываются все символы форматирования (т.е. пробелы, переводы строк, табуляции не игнорируются, как в HTML).

В общем виде ХМL-документ после пролога можно представить в следующем виде:

1. Какие должны быть имена элементов?

Имена элементов могут начинаться только с букв и символов подчеркивания и могут содержать только буквы, цифры, дефисы, точки и символы подчеркивания и не могут начинаться с сочетания «xml». В качестве имен можно использовать любые слова, зарезервированных слов нет.

1. Как вы понимаете правило соблюдения корректной вложенности?
2. <корневой>
3. <потомок>
4. <подпотомок>.....</подпотомок>
5. </потомок>
6. </корневой>

6.Какие элементы являются корневыми в XML-документе?

1. Для чего необходима валидация XML-документа?

Валидный XML документ не то же самое, что и синтаксически верный XML документ.

Первое правило для валидного XML документа то, что он должен быть синтаксически верным.

Второе правило — валидный XML документ должен соответствовать определенному типу документов.

Правила, определяющие допустимые элементы и атрибуты для XML документа, часто называются определениями документа или схемами документа.

1. В чем заключается разница между простыми и комплексными типами элементов в XML Schema?

**Элементы**, которые содержат подэлементы или имеют атрибуты, называют **элементами** **комплексного** **типа**, тогда как **элементы**, которые содержат числа (строки, даты, и т.д.), но не содержат подэлементов или атрибутов, называются **элементами** **простого** **типа**.

1. В чем заключается преимущество Schema перед DTD?

Наиболее существенным отличием между **DTD** и схемой XML является возможность создавать и использовать типы данных в схеме вместе с объявлениями элементов и атрибутов.

1. Какие комплексные типы элементов XML Schema вы знаете?

Пустые элементы

элементы, содержащие только другие элементы

элементы, содержащие только текст

элементы, содержащие другие элементы и текст

1. Какие простые типы элементов XML Schema вы знаете?

xs:string

xs:decimal

xs:integer

xs:boolean

xs:date

xs:time

1. Что означает *xs:element*?

**Xs**: **element** : определяет элемент

1. Что означает *xs:complexType*?

**xs**: complexType : **означает**, что он содержит другие элементы.

1. В чем заключается предназначение xs:sequence?

**xs**: sequence : дочерние элементы

1. Для чего предназначен *xs:attribute*?

Элемент xsl:**attribute** служит для создания узлов **атрибутов** в выходном документе. Синтаксис ¶.< xsl:**attribute** name=" имя" namespace=" пространство имен">< ! - - Содержимое: template--></ xsl:**attribute**>. **Атрибуты**: name обязательный **атрибут**, указывает имя создаваемого **атрибута**.

1. Что означает *xsi:noNamespaceSchemaLocation*?

Где находится схема

1. Для чего используется ENTITY?

ENTITY для определения сущностей в DTD с целью их использования как в связанном с DTD XML-документе, так и собственно в DTD.

1. Что такое сущности? Каким образом вы использовали сущности?

Сущность является заместителем содержания, которую можно однажды объявить и многократно использовать почти в любом месте документа. Встроенные сущности приведены в таблице 11.3

1. Какие встроенные сущности вы знаете?

Таблица 11.3

|  |  |  |
| --- | --- | --- |
| **Сущность** | **Ссылка на сущность** | **Значение** |
| lt | &lt; | < (меньше чем) |
| gt | &gt; | > (больше чем) |
| amp | &amp; | & (амперсанд) |
| apos | &apos; | ' (апостроф или одиночная кавычка) |
| quot | &quot; | " (двойная кавычка) |

1. Для чего используется ATTLIST?

ATTLIST для перечисления и объявления атрибутов, которые могут принадлежать элементу

1. Что такое валидный XML-документ? Как типы вы знаете?

Валидный XML-документ должен быть синтаксически верным и соответствовать одному из типов определения документов. Правила, определяющие допустимые элементы и атрибуты для XML документа, называются определениями документа или схемами документа.

1. Что означает следующая запись <!ENTITY name “Hello, world!”>?
2. Что означает следующая запись <!ELEMENT to (#PCDATA)>?
3. Какие параметры и значения имеет инструкция ATTLIST?

Возможные параметры:

· ***elementName*** - имя элемента, к которому относится список атрибутов.

· ***attributeName*** - имя атрибута. Этот параметр повторяется столько раз, сколько нужно для перечисления всех атрибутов, используемых с elementName.

· ***dataType*** - тип данных для атрибута, названного в параметре ***attributeName***; должен иметь одно из следующих значений:

· ***CDATA*** – атрибут содержит только символьные данные.

· ***ID*** - значение атрибута должно быть уникальным. Оно не может повторяться в других элементах или атрибутах данного документа.

1. Какие параметры имеет инструкция ELEMENT?

Возможные параметры:

· ***name*** - имя элемента. Необходимо точно воспроизвести регистр.

· ***Content*** - допустимая модель содержимого для элемента может быть одной из следующих:

· ***ANY*** — внутри элемента допускается любое содержимое. При использовании в объявлении элемента, это ключевое слово разрешает открытие неограниченной модели содержимого для элементов и всех дочерних узлов.

1. Что находится на *http://www.w3.org/2001/XMLSchema*?
2. Каким образом можно объявить DTD? Как расшифровывается аббревиатура?

|  |  |
| --- | --- |
| **Внутренняя декларация DTD** | **Внешняя декларация DTD** |
| <?xml version="1.0" encoding="UTF-8" ?>  <!DOCTYPE note [  <!ELEMENT note (to, from, Sbj, msg)>  <!ELEMENT to (#PCDATA)>  <!ELEMENT from (#PCDATA)>  <!ELEMENT Sbj (#PCDATA)>  <!ELEMENT msg (#PCDATA)>  ]> | <?xml version="1.0" encoding="UTF-8" ?>  <!DOCTYPE note SYSTEM "note.dtd">  <note>  <to>Sunny</to>  <from>Oliver</from>  <Sbj>Hello</Sbj>  <msg>This is a good day!</msg> </note> |
| <note>  <to>Sunny</to>  <from>Oliver</from>  <Sbj>Hello</Sbj>  <msg>This is a good day!</msg> </note> | **файл note.dtd содержит:**  <!ELEMENT note (to,from,heading,body)>  <!ELEMENT to (#PCDATA)>  <!ELEMENT from (#PCDATA)>  <!ELEMENT heading (#PCDATA)>  <!ELEMENT body (#PCDATA)> |