Лабораторная работа № 11 «Описание структуры XML документа средствами DTD»

**Аннотация:**Цель работы: ознакомиться с конструкциями языка XML и его реализациями, со способами описания структуры документа средствами DTD. Изучить способы обработки XML документов.

Структура XML

Первая строка XML-документа называется *объявление XML*— это необязательная строка, указывающая версию стандарта XML (обычно это 1.0), также здесь может быть указана кодировка символов и внешние зависимости.

**<?xml version="1.0" encoding="utf-8"?>**

Важнейшее обязательное синтаксическое требование заключается в том, что документ имеет только один *корневой элемент* (так же иногда называемый *элемент документа*). Это означает, что текст или другие данные всего документа должны быть расположены между единственным начальным корневым тегом и соответствующим ему конечным тегом.

<comp> компьютер</comp>

Остальная часть этого XML-документа состоит из вложенных *элементов*, некоторые из которых имеют *атрибуты* и *содержимое*. *Элемент* обычно состоит из открывающего и закрывающего тегов, обрамляющих текст и другие элементы

<video> видеокарта </video>

Кроме содержания у элемента могут быть *атрибуты* — пары имя-значение, добавляемые в открывающий тег после названия элемента. Значения атрибутов всегда заключаются в кавычки (одинарные или двойные), одно и то же имя атрибута не может встречаться дважды в одном элементе. Не рекомендуется использовать разные типы кавычек для значений атрибутов одного тега.

<video model=”6600” > видеокарта </video>

В приведённом примере у элемента «video» есть атрибут: «model», имеющий значение «6600»

В любом месте дерева может быть размещен элемент-*комментарий*. XML-комментарии размещаются внутри специального тега, начинающегося с символов <!-- и заканчивающегося символами -->. Два знака дефис (--) внутри комментария присутствовать не могут.

<!-- *Коментарий*-->

Пример:

<log>

<event date=" 27/May/1999:02:32:46 " result="success">

<ip-from> 195.151.62.18 </ip-from>

<method>GET</method>

<url-to> /misc/</url-to>

<response>200</response>

</event>

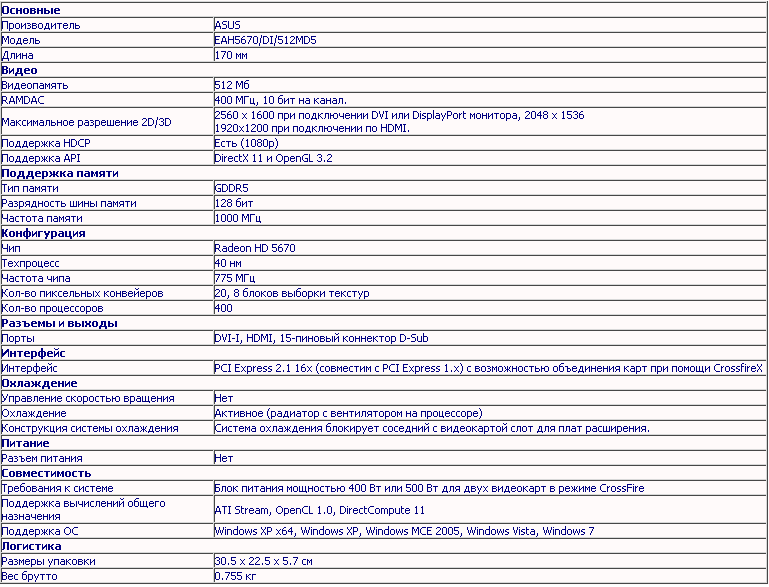
</log>

Задание:

Создайте Xml-документ на основе данной таблицы .

Использовать выделенные слова левого столбца как теги, а обычные как атрибуты.В качестве параметров используйте данные правого столбца.

Сохраните документ на личном диске и назовите его Video.xml



В результате Ваш XML-файл может выглядеть так:

<?xml version="1.0" encoding="utf-8"?>

<Harakter>

<Osnovnie proizvoditel="ASUS " model="EAH6670/DI/512MD5 " dlina=" 170 мм ">

</Osnovnie>

<Video videopamyat="512" RAMDAC="400 МГц, 10 бит на канал." HDCP="Есть (1080p)" API="DirectX 11 и OpenGL 3.2 ">

</Video>

<Pamyat tip="ddr5" shina="256" chastota="1000 МГц">

</Pamyat>

<Config chip="Radeon HD 5670" chastota\_chipa="775 МГц" kol-vo\_processorov="400 ">

</Config>

<Razemi porti="D-sub,dvi">

</Razemi>

<Interface>PCI Express 2.1</Interface>

<Ohlazhdenie upravlenie="net" ohlazhdenie="activ">

</Ohlazhdenie>

<Sovmestimost pitanie="400 Вт" os="Windows XP x64, Windows XP, Windows MCE 2005, Windows Vista, Windows 7">

</Sovmestimost>

</Harakter>

**Создание DTD-документа**

**DTD** (*Document Type Definition* определение типа документа) — включает в себя два понятия:

Термин, который используется для описания схемы документа или его части *языком схем DTD*.

* *Язык схем DTD (DTD schema language)* — искусственный язык, который используется для записи фактических синтаксических правил метаязыков разметки текста SGML и XML. С момента его внедрения другие языки схем для спецификаций, такие как XML Schema и RELAX NG, выпускаются с дополнительной функциональностью.

**Объявление элементов**

Объявления элементов образовывают перечень разрешенных названий элементов в документе, а также определяют информацию относительно тегов (являются ли они обязательными) и модели содержимого для каждого элемента.

Различные ключевые слова и символы определяют содержимое элемента:

* EMPTY — пустое содержимое
* ANY — любое содержимое
* , — указывает порядок
* | — разделение альтернатив
* () — группировка
* \* — любое количество элементов (нуль и более)
* + — по крайней мере один элемент (один и более)
* ? — необязательное наличие элемента (ноль или один)
* Если нет \*, + или ? — элемент должен быть только один

<!ELEMENT HDD - - (250|500)+>

Элемент HDD должен содержать один и более элементов 250 или 500 в произвольном порядке

#### Определение атрибутов

С каждым элементом DTD-документа можно сопоставить список атрибутов. Для этого используется директива !ATTLIST, в которой указываются имя элемента, с которым может быть сопоставлен список атрибутов и параметры каждого атрибута: его имя, тип и свойства по умолчанию.

Например:

<!ATTLIST Video model CDATA #REQUIRED>

В этом примере определен атрибут Video для элемента model. Он является обязательным.

Существуют такие типы атрибутов:

* CDATA (Character set of data) — значением атрибута могут быть любые символьные данные
* ID — значением атрибута должен быть уникальный идентификатор элемента
* IDREF — значением элемента является ссылка на элемент по его ID
* IDREFS — тоже что и IDREF, но с возможностью ссылок не по одному идентификатору, а по нескольким
* NMTOKEN — значением атрибута может быть последовательность символов, в чем-то схожая с именем (отсюда и названием — name token). Это строка, которая содержит любую комбинацию тех символов, которые разрешено использовать для имен XML.
* NMTOKENS — значением атрибута является список значений
* ENTITY — значение используется для ссылки на внешнюю сущность.
* ENTITIES — позволяет задать список внешних сущностей, разделенных пробелами.
* NOTATION — значением атрибута может быть одна из ранее определенных нотаций
* NOTATIONS — позволяет задать список нотаций.
* Listings и NOTATION-listings
* ENUMERATION — задает список возможных альтернатив значений.

Существуют такие свойства по умолчанию:

1. IMPLIED — значение атрибута указывать не обязательно;
2. REQUIRED — значение атрибута обязательно должно быть указано;
3. FIXED — значение этого атрибута задано как константа в DTD и в документе не может быть изменено;
4. некоторое конкретное значение, которое используется по умолчанию.

Пример:

<?xml encoding="koi8-r"?>

<!ELEMENT log (event)+>

<!ELEMENT event (ip-from,method,uri-to,result)>

<!ELEMENT method (#PCDATA)>

<!ELEMENT ip-from (#PCDATA)>

<!ELEMENT url-to (#PCDATA)>

<!ELEMENT response (#PCDATA)>

<!ATTLIST event

result CDATA #IMPLIED

date CDATA #IMPLIED>

Используйте созданный XML документ под названием Video.xml и убедились, что набор используемых при этом тэгов позволяет осуществлять любые манипуляции с нашей информацией. В таком случае, для того, чтобы утвердить правила нашего нового языка, т.е. список допустимых элементов, их возможное содержимое и атрибуты, мы должны создать DTD – определения.

  Создайте dtd- файл на основе следующей таблицы под именем Video.dtd и включите в XML-документ новую строчку

<!DOCTYPE VIDEO SYSTEM "Video.dtd">

В результате документ Video.dtd может выглядеть следующим образом:

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

<!-- Enter you external DTD here -->

<!ELEMENT Haracter (Osnovnie, Video,Pamyat,Config,Razemi,Interface,Ohlazhdenie,Sovmestimost)+ >

<!ELEMENT Osnovnie (#PCDATA)>

<!ATTLIST Osnovnie

proizvoditel CDATA #REQUIRED

model CDATA #REQUIRED

dlina CDATA #REQUIRED >

<!ELEMENT Video (#PCDATA)>

<!ATTLIST Video

videopamyat (256|512|1024) #REQUIRED

RAMDAC CDATA #REQUIRED

HDCP CDATA #REQUIRED

API CDATA #REQUIRED >

<!ELEMENT Pamyat (#PCDATA)>

<!ATTLIST Pamyat

tip (ddr2|ddr3|ddr5) #REQUIRED

shina (128|256|448) #REQUIRED

chastota CDATA #REQUIRED >

<!ELEMENT Config (#PCDATA)>

<!ATTLIST Config

chip CDATA #REQUIRED

chastota\_chipa CDATA #REQUIRED

kol-vo\_processorov CDATA #REQUIRED >

<!ELEMENT Razemi (#PCDATA)>

<!ATTLIST Razemi

porti CDATA #REQUIRED >

<!ELEMENT Interface (#PCDATA)>

<!ELEMENT Ohlazhdenie (#PCDATA)>

<!ATTLIST Ohlazhdenie

upravlenie (da|net) #FIXED "net"

ohlazhdenie (activ|passiv) #REQUIRED>

<!ELEMENT Sovmestimost (#PCDATA)>

<!ATTLIST Sovmestimost

pitanie CDATA #REQUIRED

os CDATA #REQUIRED>