Java.SE.13

INTEGRATION DATA FORMATS

Author: **Olga Smolyakova , PhD**
Oracle Certified Java 6 Programmer
Olga_Smolyakova@epam.com

Содержание

1. Теги, элементы, атрибуты
2. Правила XML-документа
3. Объявления XML
4. Пространства имен
5. XSD
6. DTD
7. XML parsers
8. SAX
9. STAX
10. DOM
11. JAXP
12. JDOM
13. JAXB
14. Валидация
15. Создание простого WSDL/SOAP Web-сервиса средствами Java SE

ТЕГИ, ЭЛЕМЕНТЫ, АТРИБУТЫ

Теги, элементы, атрибуты. XML

XML или **Extensible Markup Language** (Расширяемый Язык Разметки), является языком разметки, который можно использовать для создания ваших собственных тегов.

```
<note>
    <to>Вася</to>
    <from>Света</from>
    <heading>Напоминание</heading>
    <body>Позвони мне завтра!</body>
</note>
```

Теги, элементы и атрибуты

Тег - это текст между левой угловой скобкой (<) и правой угловой скобкой (>). Есть начальные теги (такие, как <name>) и конечные теги (такие, как </name>)

```
<from>
```

```
</heading>
```

Элементом является начальный тег, конечный тег и все, что есть между ними. В примере элемент <name> содержит два дочерних элемента: <title>, <first-name> и <last-name>.

```
<note>
  <to>Вася</to>
  <from>Света</from>
</note>
```

Атрибут - это пара имя-значение внутри начального тега элемента.

```
<note id="1">
```

ПРАВИЛА XML-ДОКУМЕНТА

Правила XML-документа. Корневой элемент

Корневой элемент

Документ XML должен содержаться в единственном элементе. Этот единственный элемент называется корневым элементом и содержит весь текст и любые другие элементы документа.

```
<?xml version="1.0" encoding="UTF-8"?>
<notes>
    <note id="1">
        <to>Вася</to>
        <from>Света</from>
        <heading>Напоминание</heading>
        <body>Позвони мне завтра!</body>
    </note>
</notes>
```

Правила XML-документа. Перекрытие элементов

Элементы не могут перекрываться

Элементы XML не могут перекрывать друг друга.

```
...
    <to>Вася</to>
    <from><i>Света</from>
    <heading>Напоминание</heading></i>
    <body>Позвони мне завтра!</body>
...
```

Правила XML-документа. Конечные теги

Конечные теги являются обязательными

Нельзя опускать какие-либо закрывающие теги.

```
...
        <to>Вася</to>
        <from>Света</from>
        <heading><p>Напоминание</heading>
        <body><br/>Позвони мне завтра!</body>
...
```

Правила XML-документа. Регистр

Элементы чувствительны к регистру

Элементы XML чувствительны к регистру.

```
...
    <heading>Напоминание</heading>
    <body>Позвони  мне  завтра!</BODY>
...
```

Правила XML-документа. Атрибуты

Атрибуты должны иметь значения в кавычках

Есть два правила для атрибутов в XML-документах:

- Атрибуты должны иметь значения
- Эти значения должны быть заключены в кавычки

```
<note id="1">
```

Можно использовать одинарные или двойные кавычки, но только согласованно.

Если значение атрибута содержит одинарные или двойные кавычки, можно использовать другой вид кавычек

name="Doug's car"

Также допускается сущности *"* для двойной кавычки и *&apos*; для одинарной.

Комментарии

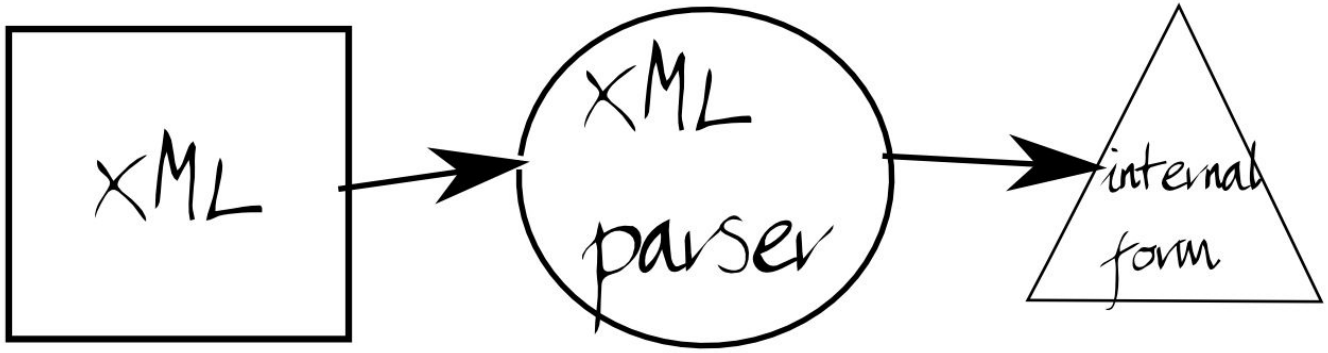
Комментарии могут появляться где угодно в документе; даже перед корневым элементом. Комментарий начинается с <!-- и заканчивается -->. Комментарий не может содержать двойного дефиса (--) нигде, кроме как в конце; за этим исключением, комментарий может содержать что угодно. Любая разметка внутри комментария игнорируется.

```
<!-- комментарий -->
```

Правила XML-документа. Парсер

Спецификация XML требует, чтобы парсер браковал любой XML-документ, который не выдерживает основные правила.

Парсер - это часть кода, которая пытается прочесть документ и интерпретировать его содержимое.



Правила XML-документа. Виды XML-документов

Есть три вида XML-документов:

- **Well-formed (синтаксически корректные)** следуют синтаксическим правилам XML.
- **Valid (валидные)** следуют синтаксическим правилам XML и соответствуют правилам, определенным в их схеме (DTD, XSD, RelaxNG или Schematron).
- **Conformance (соответствующие стандарту)** соответствует всем (в том числе изложенным человеческим языком), требованиям спецификации.

ОБЪЯВЛЕНИЯ XML

Объявления XML

Большинство XML-документов начинаются с XML-объявления, которое обеспечивает базовую информацию о документе для парсера.

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

Употребление XML-объявления рекомендуется, но не является обязательным. Если оно есть, оно должно быть первым, что есть в документе.

Объявления XML

Объявление может содержать до трех пар имя-значение (многие называют их атрибутами, хотя технически они таковыми не являются):

- **version** - используемая версия XML;
- **encoding** - набор символов, используемый в этом документе; если encoding не указан, XML-парсер предполагает набор UTF-8;
- **standalone** - может быть либо yes, либо no, определяет, может ли этот документ быть обработан без чтения каких-либо других файлов.

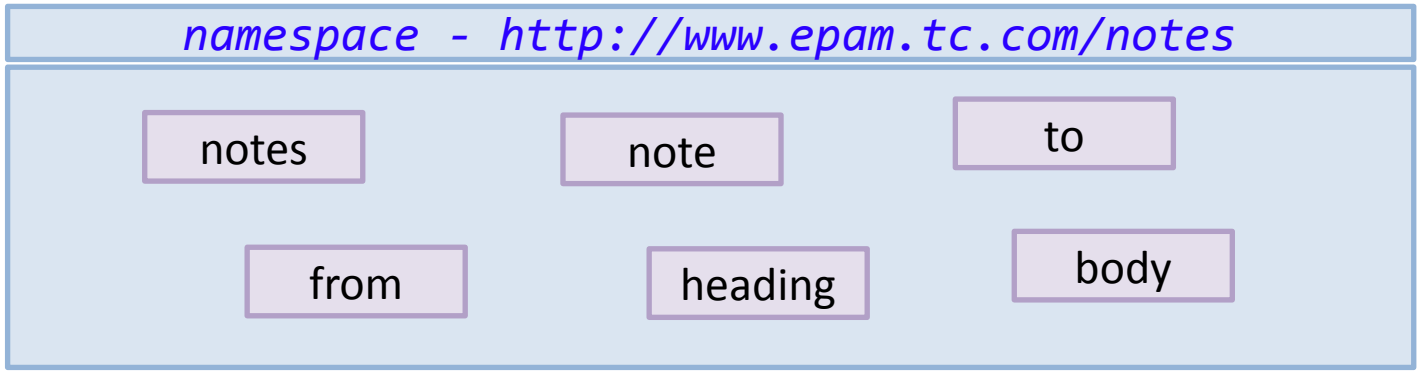
```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
```

ПРОСТРАНСТВА ИМЕН

Пространства имен. Использование пространства имен

Пространство имён (namespace) - это логическая группа уникальных идентификаторов.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<tc:notes xmlns:tc="http://www.epam.tc.com/notes"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.epam.tc.com/notes notes.xsd">

      <note id="1">
            <to>Вася</to>
            <from>Света</from>
            <heading>Напоминание</heading>
            <body>Позвони мне завтра!</body>
      </note>
</tc:notes>
```



Пространства имен. Использование пространства имен

Чтобы **использовать** пространство имен, необходимо определить префикс пространства имен и отобразить его на определенную строку. Префикс пространства имен уникален в пределах данного документа. Такое ограниченное имя (**qualified name**) однозначно идентифицирует элемент или атрибут и указывает, к какому пространству имен он относится.

```
<readers xmlns:address="http://www.xyz.com/addresses/"
     xmlns:books="http://www.zyx.com/books/"
xmlns:reader="http://www.zyx.com/readers">

     <reader:book-order>
          <address:full-address>
               <title>Mrs.</title>
               <name>Ivales</name>
          </address:full-address>
          <books:title>Lord of the Rings</books:title>
     </reader:book-order>
</readers>
```

Пространства имен. Пространство имен ≠ URL

Строка в определении пространства имен является только строкой.

Только одно важно в отношении строки пространства имен: она должна быть уникальной.

```
<readers xmlns:address="http://www.xyz.com/addresses/"
    xmlns:books="http://www.zyx.com/books/"
xmlns:reader="http://www.zyx.com/readers">
```

Пространства имен. Дочерние элементы

Определение пространства имен для определенного элемента означает, что все его дочерние элементы принадлежат к тому же пространству имен.

```
<readers xmlns:address="http://www.xyz.com/addresses/"
     xmlns:books="http://www.zyx.com/books/"
xmlns:reader="http://www.zyx.com/readers">

     <reader:book-order>
          <address:full-address>
               <title>Mrs.</title>
               <name>Ivales</name>
          </address:full-address>
          <books:title>Lord of the Rings</books:title>
     </reader:book-order>
</readers>
```

Пространства имен. Область действия пространств имен

Пространство имен действует только в пределах того элемента, атрибутом которого является его декларация.

```
<readers xmlns:reader="http://www.zyx.com/readers" >

  <reader:book-order>
    <address:full-address xmlns:address="http://www.xyz.com/addresses/">
              <title>Mrs.</title>
              <name>Ivales</name>
    </address:full-address>

    <books:title xmlns:books="http://www.zyx.com/books/">
     Lord of the Rings
    </books:title>

  </reader:book-order>
</readers>
```

Пространства имен. Правила наследования пространства имен

При использовании пространств имен важно учитывать, что атрибуты элемента не наследуют его пространство имен. Иными словами, если префикс пространства имен для атрибута не указан, то его имя не относится ни к какому пространству имен.

Пространства имен. Пространство имен по умолчанию

Существует два способа декларации пространства имен: декларация по умолчанию и явная декларация.

Декларация по умолчанию объявляет пространство имен для всех элементов и их атрибутов, которые содержатся в данном элементе.

xmlns=URI

```
<readers
  xmlns="http://www.zyx.com/readers"
  xmlns:address="http://www.xyz.com/addresses"
>

<reader:book-order>
      <address:full-address>
      <title>Mrs.</title>
      <name>Ivales</name>
</address:full-address>
</reader:book-order>

</readers>
```

Явная декларация

xmlns:

```
<readers
  xmlns:reader="http://www.zyx.com/readers"
  xmlns:address="http://www.xyz.com/addresses"
>

<reader:book-order>
      <address:full-address>
      <title>Mrs.</title>
      <name>Ivales</name>
</address:full-address>
</reader:book-order>

</readers>
```

Пространства имен. Префикс xml

Префикс xml не требует декларации. Он зарезервирован для расширений языка XML и всегда относится к пространству имен "http://www.w3.org/XML/1998/namespace".

С его помощью можно, например, задать базовый URI любого элемента с помощью атрибута xml:base следующего вида:

xml:base=URI

```
<readers xmlns:reader="http://www.zyx.com/readers"
      xmlns:book="http://www.zyx.com/books/"
      xml:base="http://www.company.com/readers/">

      <reader:book-order>

        <book:title>Lord of the Rings</book:title>
        <book:list-of-pictures xml:base="/pictures/">
           <book:book-picture xml:base="picture1.jpg">Рисунок 1
           </book:book-picture>
           <book:book-picture xml:base="picture2.jpg">Рисунок 2
           </book:book-picture>
        </book:list-of-pictures>
      </reader:book-order>
</readers>
```

XSD

XSD

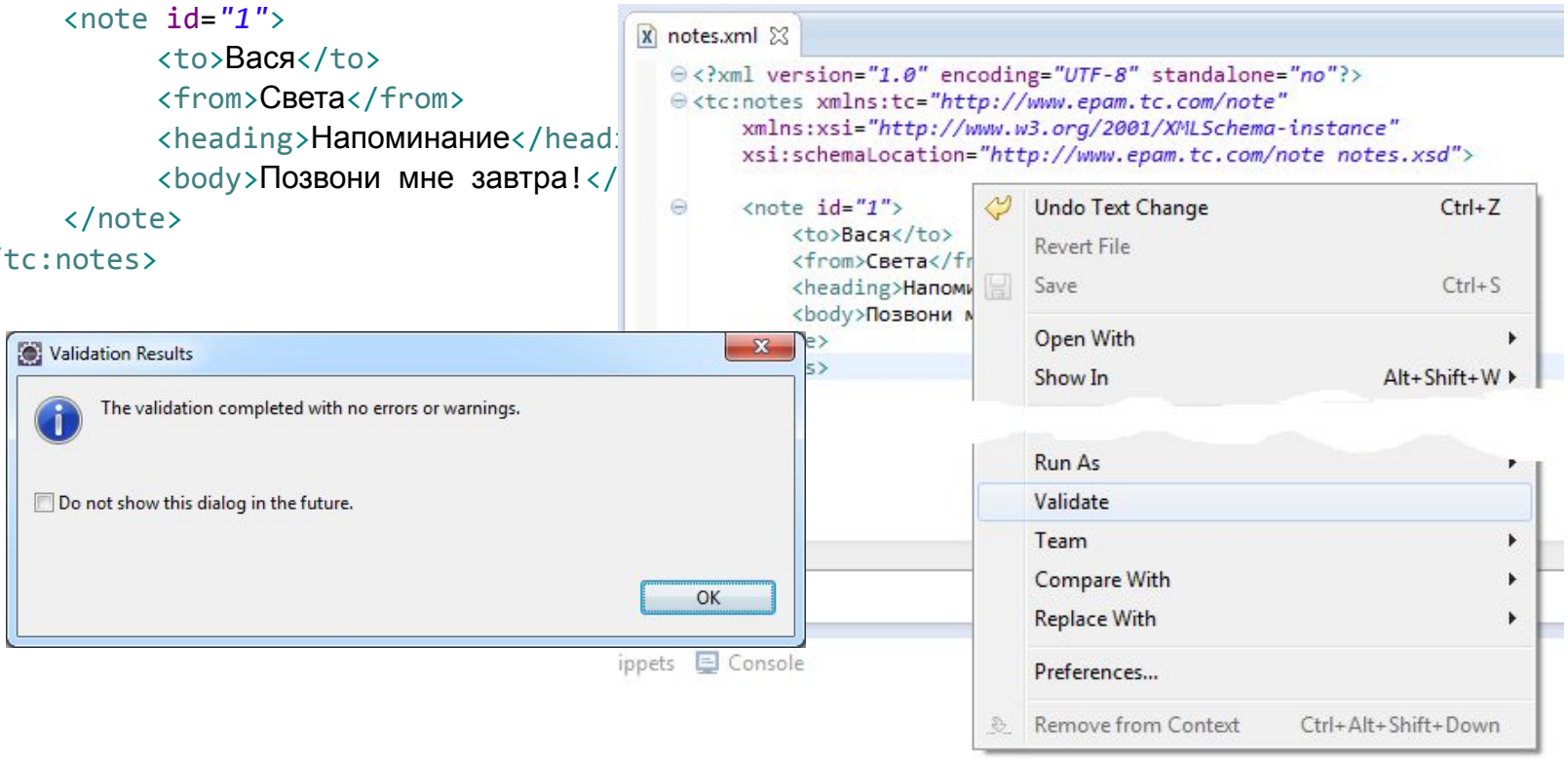
Схема XSD представляет собой строгое описание XML-документа. XSD-схема является XML-документом.

```
<?xml version="1.0" encoding="UTF-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://www.epam.tc.com/note"
xmlns:tns="http://www.epam.tc.com/note">

<element name="notes">
  <complexType>
    <sequence>
        <element name="note" type="tns:Note"
           minOccurs="1"
           maxOccurs="unbounded" />
    </sequence>
  </complexType>
</element>

<complexType name="Note">
  <sequence>
    <element name="to" type="string" />
    <element name="from" type="string" />
    <element name="heading" type="string" />
    <element name="body" type="string" />
  </sequence>
  <attribute name="id" type="int" use="required" />
</complexType>

</schema>
```

XSD. Валидация

С помощью схемы XSD можно также проверить документ на корректность.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<tc:notes xmlns:tc="http://www.epam.tc.com/note"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.epam.tc.com/note notes.xsd">

      <note id="1">
            <to>Вася</to>
            <from>Света</from>
            <heading>Напоминание</head
            <body>Позвони мне завтра!</
      </note>
</tc:notes>
```



XSD. Built-in Types

Схема XSD содержит 44 базовых типа и имеет поддержку пространств имен (**namespace**).

Built-in type	Example	Built-in type	Example
string	This is a string	**int**	-1, 126789675
normalizedString	This is a string	**unsignedInt**	0, 1267896754
token	This is a token	**long**	-1, 12678967543233
byte	-1, 126	**unsignedLong**	0, 12678967543233
unsignedByte	0, 126	**short**	-1, 12678
base64Binary	GpM7	**unsignedShort**	0, 12678
hexBinary	0FB7	**decimal**	-1.23, 0, 123.4, 1000.00
integer	-126789, -1, 0, 1, 126789	**float**	-INF, -1E4, -0, 0, 12.78E-2, 12, INF, NaN
positiveInteger	1, 126789	**double**	-INF, -1E4, -0, 0, 12.78E-2, 12, INF, NaN

XSD. Built-in Types

Built-in type	Example	Built-in type	Example
negative	-126789, -1	**boolean**	true, false 1, 0
nonNegativeInteger	0, 1, 126789	**time**	13:20:00.000, 13:20:00.000-05:00
nonPositiveInteger	-126789, -1, 0	**dateTime**	1999-05-31T13:20:00.000-05:00
duration	P1Y2M3DT10H30M12.3S	**anyURI**	http://www.example.com/,http://www.example.com/ doc.html#ID5
date	1999-05-31	**language**	en-GB, en-US, fr
gMonth	--05--	**ID**	*XML 1.0 атрибут типа ID*
gYear	1999	**IDREF**	*XML 1.0 атрибут типа IDREF*
gYearMonth	1999-02	**IDREFS**	*XML 1.0 атрибут типа IDREFS*
gDay	---31	**ENTITY**	*XML 1.0 атрибут типа ENTITY*
gMonthDay	--05-3	**ENTITIES**	*XML 1.0 атрибут типа ENTITIES*
Name	*XML 1.0 тип Name*	**NOTATION**	*XML 1.0 атрибут типа NOTATION*
QName	po:USAddress	**NMTOKEN**	*XML 1.0 атрибут типа NMTOKEN*
NCName	USAddress	**NMTOKENS**	*XML 1.0 атрибут типа NMTOKENS*

XSD. <schema>

Правила описания xsd-схемы.

<schema> - корневой элемент XML-схемы.

```
<schema>
</schema>
```

```
<schema
     xmlns:xs="http//www.w3.org/2001/XMLSchema"
xmlns="http://www.epamrd.com"
     targetNamespace="http://www.epamrd.com"
elementFormDefault="qualified"
     attributeFormDefault="qualified">
</schema>
```

XSD. <element>

<element> - основные блоки XML-документа, содержащие данные и определяющие структуру описываемого документа.

```
<element name="x" type="y"/>
```

```
<element name="customer-name" type="string" default="unknown"/>
<element name="customer-location" type="string" fixed=" UK"/>
```

Cardinality: minOccurs, maxOccurs – ограничения, накладываемые на определенные числовые элементы (по умолчанию 1).

```
<element name="Customer_order" type="integer"
        minOccurs ="0"  maxOccurs="unbounded"/>
```

```
<element name="Customer_hobbies" type="string"
         minOccurs="2"  maxOccurs="10"/>
```

XSD. Simple Types. Restriction

Simple Types (простые типы) – наследуются от встроенных типов (string, integer) и позволяют создавать собственные типы данных.

Extending Simple Types. Restriction –позволяет ограничить имеющиеся простые типы.

```
<simpleType name="LetterType">
    <restriction base="string">
        <pattern value="[a-zA-Z]"/>
    </restriction>
</simpleType>
```

XSD. Facets

Ограничения могут использовать так называемые грани (Facets).

Facet	Описание
`<minLength value="3">` `<maxLength value="8">`	Ограничение длины значений типа
`<minInclusive value="0">` `<maxInclusive value="10">`	Ограничение значений типа
`<xs:length value="30">`	Ограничение длины типа

`<enumeration value="Hippo"/>`
`<enumeration value="Zebra"/>`
`<enumeration value="Lion"/>`
	Задание значений типа в виде перечисления

`<pattern value="[0-9]"/>`	Задания ограничения значений типа в виде паттерна

Полный список возможных facets можно посмотреть в стандарте http://www.w3.org/TR/xmlschema-2/#rf-facets.

XSD. Simple Types. Union

Extending Simple Types. Union – механизм для объединения двух или более различных типов данных в один.

```
<simpleType name="SizeByNumberType">
    <restriction base="positiveInteger">

        <maxInclusive value="21"/>
    </restriction>
</simpleType>
```

```
<simpleType name="SizeByStringNameType">
    <restriction base="string">
        <enumeration value="small"/>
        <enumeration value="medium"/>
        <enumeration value="large"/>
    </restriction>
</simpleType>
```

```
<simpleType name="USClothingSizeType">
    <union memberTypes="SizeByNumberType SizeByStringNameType" />
</simpleType>
```

XSD. Simple Types. List

Extending Simple Types. List – позволяет указывать в XML ряд допустимых значений, разделенных пробелами.

```
<simpleType name="SizesinStockType">
    <list itemType="SizeByNumberType" />
</simpleType>
```

```
<xs:element name="WinningNumbers" maxOccurs="unbounded">
  <xs:simpleType>
    <xs:list>
      <xs:simpleType>
        <xs:restriction base="xs:unsignedByte">
     <xs:minInclusive value="1" />
     <xs:maxInclusive value="49" />
        </xs:restriction>
      </xs:simpleType>
    </xs:list>
  </xs:simpleType>
</xs:element>
```

```
<!-- Valid "WinningNumbers" list values -->
<WinningNumbers>  1 20 24 33 37 43
</WinningNumbers>
```

XSD. Complex Types

Complex Types (сложные типы) – это контейнеры для определения элементов, они позволяют определять дочерние элементы для других элементов.

```
<element name="Customer">
    <complexType>
        <sequence>
            <element name="Dob" type="date" />
            <element name="Address" type="string" />
        </sequence>
    </complexType>
</element>
```

Compositors – определяет правила описания дочерних элементов в родительском в документе XML.

sequence	Дочерние элементы, описываемые xsd-схемой, могут появляться в XML-документе только в указанном порядке.
choice	Только один из дочерних элементов, описываемых xsd-схемой, может появиться в XML-документе.
all	Дочерние элементы, описываемые xsd-схемой, могут появляться в XML-документе в любом порядке.

XSD. Global Types

Global Types – сложные типы данных можно объявить не только внутри элемента, но и вне его.

```
<complexType name="AddressType">
    <sequence>
        <element name="Line1" type="string"/>

        <element name="Line2" type="string"/>
    </sequence>
</complexType>
```

```
<element name="Customer">
    <complexType>
        <sequence>
            <element name="Dob" type="date"/>
            <element name="Address" type="xs:AddressType"/>
        </sequence>
    </complexType>
</element>
```

XSD. Attributes

Attributes – атрибуты предоставляют дополнительную информацию в пределах элемента.

```
<attribute name="x" type="y"/>
```

```
<attribute name="ID" type="string"/>
<attribute name="ID" type="string" use="optional"/>
<attribute name="ID" type="string" use=" prohibited"/>
```

XSD. Mixed Content

Mixed Content – позволяет смешивать элементы и данные.

```
<element   name="MarkedUpDesc">
    <complexType   mixed="true">
        <sequence>
    <element   name="Bold"   type="string"   />
    <element   name="Italic"   type="string"   />
        </sequence>
    </complexType>
</element>
```

```
<MarkedUpDesc>
    This is an
    <Bold>Example</Bold>
    of
    <Italic>Mixed</Italic>
    Content,
    Note there are elements mixed in with the elements data.
</MarkedUpDesc>
```

XSD. <group>, <attributeGroup>

<group> и **<attributeGroup>** - объединяют элементы(атрибуты) в группы, позволяя на них ссылаться. Такие группы не могут быть расширены или ограничены.

```
<group name="CustomerDataGroup">
  <sequence>
    <element name="Forename" type="string" />
    <element name="Surname" type="string" />
    <element name="Dob" type="date" />
  </sequence>
</group>
<attributeGroup name="DobPropertiesGroup">
  <attribute name="Day" type="string" />
  <attribute name="Month" type="string" />
  <attribute name="Year" type="integer" />
</attributeGroup>
```

```
<complexType name="Customer">
    <sequence>
        <group ref="u:CustomerDataGroup"/>
        <element name="..." type="..."/>
    </sequence>
    <attributeGroup ref="u:DobPropertiesGroup"/>
</complexType>
```

XSD. <any>

<any> - определяет, что документ может содержать элементы, неопределенные в XML-схеме.

```
<element name="Message">
  <complexType>
    <sequence>
      <element name="DateSent" type="date" />
      <element name="Sender" type="string" />
      <element name="Content">
        <complexType>
          <sequence>
            <any />
          </sequence>
        </complexType>
      </element>
    </sequence>
  </complexType>
</element>
```

```
<Message>
 <DateSent>2000-01-12</DateSent>
 <Sender>Admin</Sender>
 <Content>
  <AccountCreationRequest>
   <AccountName>Fred</AccountName>
  </AccountCreationRequest>
 </Content>
</Message>
```

XSD. <anyAttribute>

<anyAttribute> - позволяет указывать в элементе атрибут, не определенный в XML-схеме.

```
<element name="Sender">
  <complexType>
    <simpleContent>
      <extension base="string">
        <anyAttribute />
      </extension>
    </simpleContent>
  </complexType>
</element>
```

```
<Sender ID="7687">Fred</Sender
>
```

XSD. Квалификация.

Квалификация

Квалификация необходима для однозначного разграничения пространств имен элементов и атрибутов.

Для того, чтобы все локально объявленные элементы были квалифицированы, необходимо установить значение **elementFormDefault** (<scheme>) равным **qualified**.

Атрибуты, которые должны быть квалифицированы (либо потому что объявлены глобально, либо потому что признак **attributeFormDefault**, установлен в **qualified**), в документах появляются с префиксом.

```
<xs:schema
     xmlns:xs="http//www.w3.org/2001/XMLSchema"
     xmlns="http://www.epam.com"
     targetNamespace="http://www.epam.com"
     elementFormDefault="qualified"
     attributeFormDefault="qualified">
</schema>
```

XSD. Квалификация.

Card.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www.epamrd.org/card"
        xmlns:tns="http://www.epamrd.org/card"
        elementFormDefault="qualified">


    <complexType name="CardType">
        <sequence>
            <element name="message" type="string" />
            <element name="color" type="string" />
        </sequence>
    </complexType>

</schema>
```

XSD. Квалификация.

Sock.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www.epamrd.org/sock"
        xmlns:tns="http://www.epamrd.org/sock"
        elementFormDefault="qualified"
        attributeFormDefault="qualified">

    <complexType name="SockType">
        <sequence>
            <element name="color" type="string"/>
        </sequence>
        <attribute name="size" type="integer" />
    </complexType>

</schema>
```

XSD. Квалификация.

Presents.xs

```
<?xml version="1.0" encoding="UTF-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www.epamrd.org/Presents"
        xmlns:tns="http://www.epamrd.org/Presents"
        xmlns:s="http://www.epamrd.org/sock"
        xmlns:c="http://www.epamrd.org/card"
        elementFormDefault="qualified">
  <import schemaLocation="Sock.xsd" namespace="http://www.epamrd.org/sock" />
  <import schemaLocation="Card.xsd" namespace="http://www.epamrd.org/card" />

  <element name="presents" type="tns:PresentType" />

  <complexType name="PresentType">
    <sequence>
      <element name="present">
        <complexType>
          <sequence>
     <element name="sock" type="s:SockType" />
     <element name="card" type="c:CardType" />
          </sequence>
        </complexType>
      </element>
    </sequence>
  </complexType>
</schema>
```

XSD. Квалификация.

Presents.xm
l

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

<pr:presents xmlns:pr="http://www.epamrd.org/Presents"
   xmlns:s="http://www.epamrd.org/sock"
   xmlns:c="http://www.epamrd.org/card"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.epamrd.org/Presents Presents.xsd">

    <pr:present>
        <pr:sock s:size="4">
            <s:color>red</s:color>
        </pr:sock>
        <pr:card>
            <c:message>Happy New Year!</c:message>
            <c:color>gold</c:color>
        </pr:card>
    </pr:present>

</pr:presents>
```

DTD

DTD. Document Type Definition

Для описания структуры XML-документа используется язык описания **DTD (Document Type Definition)**.

DTD определяет, какие теги (элементы) могут использоваться в XML-документе, как эти элементы связаны между собой (например, указывать на то, что элемент **<student>** включает дочерние элементы **<name>**, **<telephone>** и **<address>**), какие атрибуты имеет тот или иной элемент.

DTD. Формирование DTD

Подключение DTD

1.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<! DOCTYPE students SYSTEM "students.dtd">
```

2.

```
<?xml version="1.0" ?>
<! DOCTYPE student
[<!ELEMENT student (name, telephone, address)><!--далее
идет описание элементов name, telephone, address -->]
>
```

DTD. Пример

students.xml

```
<?xml version="1.0" ?>
<!DOCTYPE students SYSTEM "students.dtd" >
<students>
  <student login="mit"
    <name>Mitar Alex</
    <telephone>2456474
    <address>
      <country>Belarus
      <city>Minsk</city
      <street>Kalinovs
    </address>
  </student>
  <student login="pus"
    <name>Pashkun Alex
      <telephone>34537
      <address>
        <country>Belar
        <city>Brest</ci
        <street>Knorin
        </address>
  </student>
</students>
```

students.dtd

```
<!ELEMENT students (student)+>
<!ELEMENT student (name, telephone, address)>
<!ATTLIST student
    login ID #REQUIRED
    faculty CDATA #REQUIRED
>
<!ELEMENT name (#PCDATA)>
<!ELEMENT telephone (#PCDATA)>
<!ELEMENT address (country, city, street)>
<!ELEMENT country (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT street (#PCDATA)>
```

DTD. <!ELEMENT>

Описание элемента

```
<!ELEMENT name (#PCDATA)>
<!ELEMENT telephone (#PCDATA)>
<!ELEMENT address (country, city, street)>
```

PCDATA. - элементы могут содержать любую информацию, с которой может работать программа-анализатор (**PCDATA** – parsed character data). Есть также маркеры **EMPTY** – элемент пуст и **ANY** – содержимое специально не описывается.

Если в определении элемента указывается "смешанное" содержимое, т.е. текстовые данные или набор элементов, то необходимо сначала указать PCDATA, а затем разделенный символом "|" список элементов.

DTD. <!ELEMENT>

Для того, чтобы указать количество повторений включений элементов могут использоваться символы: ‘**+**’(один или много), ‘*****’(0 или много), ‘**?**’(0 или 1)

- **<!ELEMENT student (name, telephone, address)>** - элемент **student** содержит один и только один элемент **name**, **telephone** и **address**.

Если существует несколько вариантов содержимого элементов, то используется символ ‘|’ (или).

- **<!ELEMENT student (#PCDATA | body)> -** элемент **student** может содержать либо дочерний элемент **body**, либо **PCDATA**.
- **<!ELEMENT issue (title, author+, table-of-contents?)>** - внутри.
- **<!ELEMENT flower (PCDATA | title)*>**

DTD. <!ATTLIST>

Описание атрибутов

```
<!ATTLIST название_елемента название_атрибута
тип_атрибута значение_по_умолчанию >
```

Например:

```
<!ATTLIST student
login ID #REQUIRED
faculty CDATA #REQUIRED>
```

DTD. <!ATTLIST>

Существует несколько возможных значений атрибута, это:

- **CDATA** – значением атрибута является любая последовательность символов;
- **ID** – определяет уникальный идентификатор элемента в документе;
- **IDREF** (**IDREFS**) – значением атрибута будет идентификатор (список идентификаторов), определенный в документе;
- **ENTITY** (**ENTITIES**) – содержит имя внешней сущности (несколько имен, разделенных запятыми);
- **NMTOKEN** (**NMTOKENS**) – слово (несколько слов, разделенных пробелами).

DTD. <!ATTLIST>

Опционально можно задать значение по умолчанию для каждого атрибута. Значения по умолчанию могут быть следующими:

- **#REQUIRED** – означает, что атрибут должен присутствовать в элементе;
- **#IMPLIED** – означает, что атрибут может отсутствовать, и если указано значение по умолчанию, то анализатор подставит его.
- **#FIXED defaultValue** – означает, что атрибут может принимать лишь одно значение, то, которое указано в DTD.

defaultValue – значение по умолчанию, устанавливаемое парсером при отсутствии атрибута. Если атрибут имеет параметр **#FIXED**, то за ним должно следовать **defaultValue**.

Если в документе атрибуту не будет присвоено никакого значения, то его значение будет равно заданному в DTD.

DTD. <!ATTLIST>

defaultValue – значение по умолчанию, устанавливаемое парсером при отсутствии атрибута. Если атрибут имеет параметр **#FIXED**, то за ним должно следовать **defaultValue**.

Если в документе атрибуту не будет присвоено никакого значения, то его значение будет равно заданному в DTD. Значение атрибута всегда должно указываться в кавычках.

DTD. <!ENTITY>

Определение сущности

Сущность (entity) представляет собой некоторое определение, чье содержимое может быть повторно использовано в документе. Описывается сущность с помощью дескриптора **!ENTITY**:

```
<!ENTITY company 'Sun Microsystems'>

<sender>&company;</sender>
```

Программа-анализатор, которая будет обрабатывать файл, автоматически подставит значение Sun Microsystems вместо **&company**.

DTD. <!ENTITY>

Для повторного использования содержимого внутри описания DTD используются параметрические (параметризованные) сущности.

```
<!ENTITY  % elementGroup "firstName,
lastName,gender, address, phone">

<!ELEMENT employee (%elementGroup)>
<!ELEMENT contact (%elementGroup)>
```

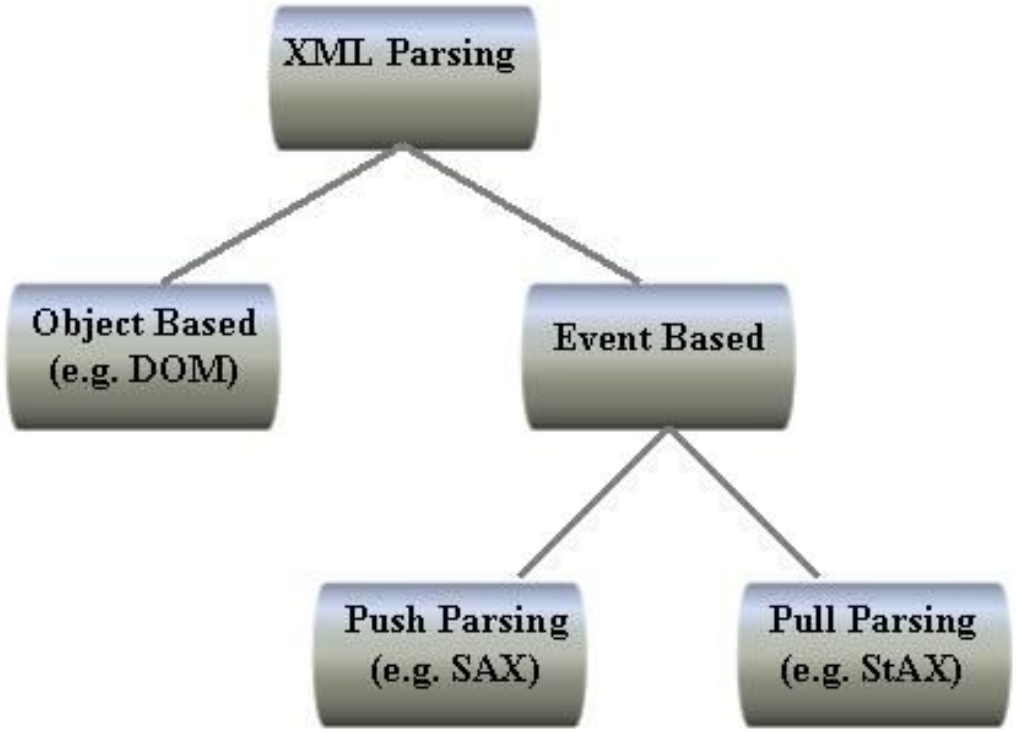
DTD. <!ENTITY>

В XML включены внутренние определения для символов. Кроме этого, есть внешние определения, которые позволяют включать содержимое внешнего файла:

```
<!ENTITY logotype SYSTEM "/image.gif" NDATA GIF87A>
```

XML PARSERS

XML Parsers



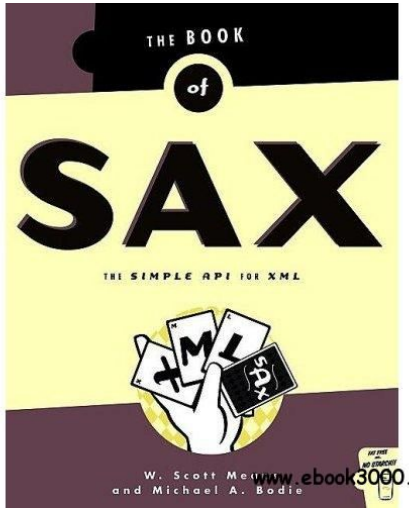
SAX

SAX. Введение

SAX - это событийный парсер для XML, т.е. он последовательно читает и разбирает данные из входного потока (это может быть файл, сетевое соединение, или любой другой InputStream).

Когда парсер находит структурный элемент (открывающий тег, закрывающий тег, и т.п.), он оповещает об этом слушателя (обработчик события), и передает ему в качестве параметра найденный элемент.

SAX делает возможным привязку специфичного для приложения кода к событиям.



SAX. XMLReader

SAX (SAX2) определяет интерфейс,

org.xml.sax.XMLReader

который должны реализовывать все SAX-совместимые анализаторы XML. (Благодаря этому SAX точно знает, какие методы доступны для обратного вызова и использования в приложении).

```
// создание экземпляра класса Reader
org.xml.sax.XMLReader reader =
       XMLReaderFactory.createXMLReader();

// делаем что-то с помощью анализатора
reader.parse(url);
```

SAX. InputSource

Для анализа документа применяется метод **parse()** класса org.xml.sax.XMLReader.

У качестве параметра может выступать экземпляр класса **org.xml.sax.InputSource**, либо строка, содержащая URI.

```
// создание экземпляра класса Reader
org.xml.sax.XMLReader reader =

XMLReaderFactory.createXMLReader(vendorParserClass);

// регистрируем обработчик содержимого
// регистрируем обработчик ошибок
// анализируем
InputSource inputSource = new InputSource(xmlURI);
reader.parse(inputSource);
```

SAX. InputSource

Используя **InputSource** и заключив в него переданный URI можно определить системный идентификатор документа. По сути дела, устанавливается путь к документу для анализатора, что и позволяет разрешать все относительные пути внутри этого документа.

SAX гарантирует, что анализатор никогда не изменит объект InputSource, передаваемый в качестве аргумента методу parse().

```
InputSource inputSource =
    new InputSource(
        new java.io.FileInputStream(
            new java.io.File(xmlURI)));
```

Handler

Обработчик содержимого - это набор методов обратного вызова SAX, позволяющих программистам связывать код приложения с событиями, возникающими во время синтаксического разбора документа.

SAX обрабатывает документ последовательно и не загружает его в память целиком.

SAX. Handler

В SAX 2.0 определены четыре основных интерфейса-обработчика:

- **org.xml.sax.ContentHandler** –обработчик событий документа
- **org.xml.sax.ErrorHandler** – обработки ошибочных ситуаций
- **org.xml.sax.DTDHandler** – обработчик событий при анализе DTD-описаний
- **org.xml.sax.EntityResolver** - обработчик событий загрузки DTD-описаний (создан специально для интерпретации внешних сущностей, на которые ссылается XML-документ)

Классы, реализующие эти интерфейсы можно зарегистрировать в анализаторе с помощью методов **setContentHandler(); setEntityResolver(); setDTDHandler(); setErrorHandler();**

SAX. ContentHandler

```
public interface ContentHandler {

   void setDocumentLocator(Locator locator);
   void startDocument() throws SAXException;
   void endDocument() throws SAXException;
   void startPrefixMapping(String prefix, String uri) throws SAXException;
   void endPrefixMapping(String prefix) throws SAXException;

   void startElement(String uri, String localName, String qName, Attributes atts) throws
                                                                        SAXException;

   void endElement(String uri, String localName, String qName) throws SAXException;
   void characters(char ch[], int start, int length) throws SAXException;
   void ignorableWhitespace(char ch[], int start, int length) throws SAXException;
   void processingInstruction(String target, String data) throws SAXException;
   void skippedEntity(String name) throws SAXException;
}
```

SAX. ContentHandler

Методы интерфейса ContentHandler

Метод	Назначение
setDocumentLocator(Locator locator)	Указатель позиции в документе, действителен только для текущего цикла анализа.
startDocument()	Вызывается первым во всем процессе анализа документа, за исключением метода setDocumentLocator()
endDocument()	Вызывается последним, в том числе и среди методов остальных обработчиков SAX
startElement(String uri, String localName, String qName, Attributes atts)	Сообщает о начале анализа документа, предоставляет приложению информацию об элементе XMLи любых его атрибутах
endElement(String uri, String localName, String qName)	Сообщает о достижении закрывающего тега элемента.

SAX. ContentHandler

Методы интерфейса ContentHandler

Метод	Назначение
startPrefixMapping(String prefix, String uri)	Вызывается перед методом, связанным с элементом, в котором пространство имен объявлено.
endPrefixMapping(String prefix)	Вызывается после закрытия элемента в котором пространство имен объявлено.

В SAX 2 поддержка пространства имен осуществляется на уровне элементов. Это позволяет различать пространство имен элемента, представленное префиксом элемента и связанным с этим префиксом URI, и локальное имя элемента (имя без префикса).

Область отображения префикса – это элемент с атрибутом xmlns, объявляющий пространство имен.

SAX. ContentHandler

Методы интерфейса ContentHandler

Метод	Назначение
characters(**char** ch[], **int** start, **int** length)	Сообщает о достижении символьного значения элемента
ignorableWhitespace(**char** ch[], **int** start, **int** length)	Метод обратного вызова для необрабатываемых символов между элементами; должен вызываться только при наличии DTD или XML схемы, задающих ограничения.
processingInstruction(String target, String data)	Метод обратного вызова для обработки PI-инструкций
skippedEntity(String name)	Метод обратного вызова, который выполняется, если анализатор без проверки действительности пропускает сущность (анализаторы без проверки действительности не обязаны (но могут) интерпретировать ссылки на сущности).

SAX. DocumentHandler

Интерфейс DocumentHandler

```
package org.xml.sax;
/**
 * @deprecated This interface has been replaced by the SAX2
 *             {@link org.xml.sax.ContentHandler ContentHandler}
 *             interface, which includes Namespace support.
 */
public interface DocumentHandler {
   void setDocumentLocator(Locator locator);
   void startDocument() throws SAXException;
   void endDocument() throws SAXException;
   void startElement(String name, AttributeList atts) throws SAXException;
   void endElement(String name) throws SAXException;
   void characters(char ch[], int start, int length) throws SAXException;
   void ignorableWhitespace(char ch[], int start, int length) throws SAXException;
   void processingInstruction(String target, String data) throws SAXException;
}
```

SAX. ErrorHandler

```
package org.xml.sax;

/**
 * Basic interface for SAX error handlers.
 */
public interface ErrorHandler {

   /** Предупреждения. */
   void warning (SAXParseException exception) throws SAXException;

   /** Некритические ошибки. */
   void error (SAXParseException exception) throws SAXException;

   /** Критические ошибки. */
   void fatalError (SAXParseException exception) throws SAXException;

}
```

SAX. DTDHandler

Интерфейс DTDHandler

```
public interface DTDHandler {
  public abstract void notationDecl(String name, String publicId,
             String systemId)  throws SAXException;
  public abstract void unparsedEntityDecl(String name, String
publicId,
     String systemId, String notationName)  throws SAXException;
}
```

Интерфейс DTDHandler позволяет получать уведомление, когда анализатор встречает неанализируемую сущность или объявление нотации.

SAX. EntityResolver

Интерфейс EntityResolver

```
public interface EntityResolver {

  public abstract InputSource resolveEntity(String publicId,
    String systemId) throws SAXException, IOException;

}
```

Интерфейс интерпретирует сущности. Обычно сущность по публичному или системному идентификатору интерпретируется анализатором XML. Если метод resolveEntity() возвращает null, этот процесс протекает в традиционном варианте. Если вернуть из метода корректный объект InputSource, то вместо указанного публичного или системного идентификатора в качестве значения ссылки на сущности будет использоваться этот объект.

SAX. Locator, DefaultHandler

```
public interface Locator {
    public abstract String getPublicId();
    public abstract String getSystemId();
    public abstract int getLineNumber();
    public abstract int getColumnNumber();
}
```

Интерфейс Locator позволяет определить текущую позицию в XML файле. Поскольку эта позиция действительна только для текущего цикла анализа, локатор следует использовать только в области видимости реализации интерфейса ContentHandler.

```
public class DefaultHandler
implements EntityResolver, DTDHandler, ContentHandler, ErrorHandler
{
…
}
```

Класс DefaultHandler реализует интерфейсы ContentHandler, ErrorHandler, EntityResolver, DTDHandler и предоставляет пустые реализации для каждого метода каждого интерфейса.

SAX. Пример анализа xml-документа

menu.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<breakfast-menu>
  <food id="1">
    <name>Belgian Waffles</name>
    <price>$5.95</price>
    <description>
      two of our famous Belgian Waffles with plenty of real maple syrup
    </description>
    <calories>650</calories>
  </food>
  <food id="2">
    <name>Strawberry Belgian Waffles</name>
    <price>$7.95</price>
    <description>
      light Belgian waffles covered with strawberrys and whipped cream
    </description>
    <calories>900</calories>
  </food>
     ...
</breakfast-menu>
```

SAX. Пример анализа xml-документа

```
public enum MenuTagName {
    NAME, PRICE, DESCRIPTION, CALORIES, FOOD, BREAKFAST_MENU
}
```

```
public class MenuSaxHandler extends DefaultHandler {
  private List<Food> foodList = new ArrayList<Food>();
  private Food food;
  private StringBuilder text;

  public List<Food> getFoodList() {
    return foodList;
  }

  public void startDocument() throws SAXException {
    System.out.println("Parsing started.");
  }

  public void endDocument() throws SAXException {
    System.out.println("Parsing ended.");
  }
```

SAX. Пример анализа xml-документа

```
  public void startElement(String uri, String localName, String
qName,
              Attributes attributes)  throws SAXException {
    System.out.println("startElement -> " + "uri: "
     + uri + ", localName: " + localName + ", qName: " + qName);

     text = new StringBuilder();
     if (qName.equals("food")){
     food = new Food();
              food.setId((Integer.parseInt(attributes.getValue("id")))
);
    }
   }

  public void characters(char[] buffer, int start, int length) {
     text.append(buffer, start, length);
   }
```

SAX. Пример анализа xml-документа

```
public void endElement(String uri, String localName, String qName)
            throws SAXException {
   MenuTagName tagName =
    MenuTagName.valueOf(qName.toUpperCase().replace( "-", "_"));
   switch(tagName){
    case NAME:
         food.setName(text.toString());    break;
    case PRICE:
         food.setPrice(text.toString());   break;
    case DESCRIPTION:
         food.setDescription(text.toString()); break;
    case CALORIES:
         food.setCalories(Integer.parseInt(text.toString()));
             break;
    case FOOD:
         foodList.add(food);
         food = null;
         break;
    }
}
```

SAX. Пример анализа xml-документа

```
  public void warning(SAXParseException exception) {
    System.err.println("WARNING: line " + exception.getLineNumber()
+ ": "   + exception.getMessage());
  }

  public void error(SAXParseException exception) {
    System.err.println("ERROR: line " + exception.getLineNumber() +
": "+ exception.getMessage());
  }

  public void fatalError(SAXParseException exception)  throws
SAXException {
        System.err.println("FATAL: line " +
exception.getLineNumber() +  ": "
                + exception.getMessage());
        throw (exception);
    }
}
```

SAX. Пример анализа xml-документа

```
public class Food {
    private int id;
    private String name;
    private String price;
    private String description;
    private int calories;
}
```

SAX. Пример анализа xml-документа

```
import org.xml.sax.InputSource;
import org.xml.sax.SAXException;
import org.xml.sax.XMLReader;
import org.xml.sax.helpers.XMLReaderFactory;

public class SaxDemo {

    public static void main(String[] args)  throws
ParserConfigurationException, SAXException, IOException {

        XMLReader reader = XMLReaderFactory. createXMLReader();
        MenuSaxHandler handler =  new MenuSaxHandler();
        reader.setContentHandler(handler);
        reader.parse(new InputSource("menu.xml"));

        List<Food> menu = handler.getFoodList();

    }
}
```

SAX. Расширенные возможности SAX2

Расширенные возможности SAX 2

```
public interface XMLReader
{
    boolean getFeature (String name)
        throws SAXNotRecognizedException, SAXNotSupportedException;
    void setFeature (String name, boolean value)
        throws SAXNotRecognizedException, SAXNotSupportedException;
    Object getProperty (String name)
        throws SAXNotRecognizedException, SAXNotSupportedException;
    void setProperty (String name, Object value)
        throws SAXNotRecognizedException, SAXNotSupportedException;

    // Event handlers.
}
```

В SAX 2 определен стандартный механизм для установки свойств и возможностей анализатора, что позволяет добавлять новые свойства и возможности, если они утверждены консорциумом W3C, без использования фирменных расширений и методов.

SAX. Расширенные возможности SAX2

```
// включение проверки действительности
reader.setFeature("http://xml.org/sax/features/validation" ,  true);

// включение обработки пространств имен
reader.setFeature("http://xml.org/sax/features/namespaces" , true);

// включение канонизации строк
reader.setFeature("http://xml.org/sax/features/string-interning" ,
true);

// отключение обработки схем
reader.setFeature("http://apache.org/xml/features/validation/schema
", false);
```

На страницах

http://xerces.apache.org/xerces-j/features.html

http://xerces.apache.org/xerces-j/properties.html

перечислены все возможности и свойства, поддерживаемые анализатором Apache Xerces.

SAX. org.xml.sax.XMLFilter

org.xml.sax.XMLFilter

```
public interface XMLFilter extends XMLReader
{
    void setParent (XMLReader parent);
    XMLReader getParent ();
}
```

XMLFilter предназначен для создания цепей реализаций XMLReader посредством фильтрации.

SAX. XMLFilter

org.xml.sax.helpers.XMLFilterImpl

```
public class XMLFilterImpl implements XMLFilter, EntityResolver,
DTDHandler, ContentHandler, ErrorHandler {
    // Construct an empty XML filter, with no parent.
    public XMLFilterImpl() {
        super();
    }
    // Construct an XML filter with the specified parent.
    public XMLFilterImpl(XMLReader parent) {
        super();
        setParent(parent);
    }
}
```

XMLFilterImpl позволяет создать конвейер для всех событий SAX, при этом переопределив любые методы, которые необходимо переопределить для добавления в конвейер поведения, определяемого разработчиком приложения.

XMLFilterImpl передаст в неизменном виде события, для которых не были переопределены методы.

SAX. XMLFilter, пример

```
public class NamespaceFilter  extends XMLFilterImpl {

  public NamespaceFilter(XMLReader reader){
     super(reader);
  }

  public void startPrefixMapping(String prefix, String uri)
              throws SAXException {
    System.out.println("startPrefixMapping in NamespaceFilter -"  +
          prefix + ", " + uri);
    super.startPrefixMapping(prefix, uri+ "2");
  }
}
```

SAX. XMLFilter, пример

```
public class ElementFilter extends XMLFilterImpl{

  public void startElement(String uri, String localName, String qName,
              Attributes attributes)  throws SAXException {

    System.out.println("startElement in ElementFilter" );
    super.startElement(uri+"2", localName, qName, attributes);
  }
}
```

SAX. XMLFilter, пример

```
public class SAXFilterDemo {

public static void main(String[] args)
   throws ParserConfigurationException, SAXException, IOException {

    XMLReader reader = XMLReaderFactory. createXMLReader();
    NamespaceFilter namespaceFilter =  new NamespaceFilter(reader);
    ElementFilter elementFilter =  new ElementFilter();
    elementFilter.setParent(namespaceFilter);
    MenuSaxHandler handler =  new MenuSaxHandler();
    elementFilter.setContentHandler(handler);
    elementFilter.parse( new InputSource("menu.xml"));
}

}
```

Streaming API for XML (StAX)
JSR (Java Specification Request) №173

https://jcp.org/en/jsr/detail?id=173

StAX. Введение

StAX (Streaming API for XML), который еще называют pull-парсером, включен в JDK, начиная с версии Java SE 6.

Он похож на SAX отсутствием объектной модели в памяти и последовательным продвижением по XML, но в StAX не требуется реализация интерфейсов, и приложение само “командует” StAX-парсеру переход к следующему элементу XML.

Кроме того, в отличие от SAX, данный парсер предлагает API для создания XML-документа.

StAX. Типы API

Работая со StAX можно использовать два типа API

- Iterator API (удобное и простое в использовании)
- Cursor API (быстрое, но низкоуровневое)

Iterator API	Cursor API
XMLEvent XMLEventReader XMLEventWriter	XmlStreamReader XMLStreamWriter

StAX. Cursor API

Основными классами StAX Cursor API являются

- javax.xml.stream.**XMLInputFactory**,
- javax.xml.stream.**XMLStreamReader**

и

- javax.xml.stream.**XMLOutputFactory**,
- javax.xml.stream.**XMLStreamWriter**,

которые соответственно используются для чтения и создания XML-документа.

StAX. XMLStreamReader

Для чтения XML надо получить ссылку на **XMLStreamReader**:

```
StringReader stringReader = new StringReader(xmlString);
XMLInputFactory inputFactory = XMLInputFactory.newInstance();
XMLStreamReader reader =
inputFactory.createXMLStreamReader(stringReader);
```

после чего **XMLStreamReader** можно применять аналогично интерфейсу **Iterator**, используя методы **hasNext()** и **next()**:

- **boolean hasNext()** – показывает, есть ли еще элементы;
- **int next()** – переходит к следующей вершине XML, возвращая ее тип.

StAX. XMLStreamConstants

Возможные типы вершин:

```
public interface XMLStreamConstants {
    int START_ELEMENT = 1;
    int END_ELEMENT = 2;
    int PROCESSING_INSTRUCTION = 3;
    int CHARACTERS = 4;
    int COMMENT = 5;
    int SPACE = 6;
    int START_DOCUMENT = 7;
    int END_DOCUMENT = 8;
    int ENTITY_REFERENCE = 9;
    int ATTRIBUTE = 10;
    int DTD = 11;
    int CDATA = 12;
    int NAMESPACE = 13;
    int NOTATION_DECLARATION = 14;
    int ENTITY_DECLARATION = 15;
}
```

StAX. Использование

Далее данные извлекаются применением методов:

- **String getLocalName()** – возвращает название тега;
- **String getAttributeValue(NAMESPACE_URI, ATTRIBUTE_NAME)** – возвращает значение атрибута;
- **String getText()** – возвращает текст тега.

StAX. Пример анализа

```
public enum MenuTagName {
     NAME, PRICE, DESCRIPTION, CALORIES, FOOD, BREAKFAST_MENU;

public static MenuTagName getElementTagName(String element) {
     switch (element) {
     case "food": return FOOD;
     case "price":  return PRICE;
     case "description": return DESCRIPTION;
     case "calories": return CALORIES;
     case "breakfast-menu": return BREAKFAST_MENU;
     case "name": return NAME;
     default:
     throw new EnumConstantNotPresentException(MenuTagName. class,
                     element);
     }
}
}
```

StAX. Пример анализа

```
public class StAXMenuParser {

  public static void main(String[] args)  throws FileNotFoundException {
    XMLInputFactory inputFactory = XMLInputFactory. newInstance();
    try {
     InputStream input =  new FileInputStream("menu.xml");

     XMLStreamReader reader =
         inputFactory.createXMLStreamReader(input);
     List<Food> menu =  process(reader);

     for (Food food : menu) {
         System.out.println(food.getName());
         System.out.println(food.getCalories());
     }
  } catch (XMLStreamException e) {
    e.printStackTrace();
  }
  }
```

StAX. Пример анализа

```
private static List<Food> process(XMLStreamReader reader)
              throws XMLStreamException {
  List<Food> menu = new ArrayList<Food>();
  Food food = null;
  MenuTagName elementName = null;
  while (reader.hasNext()) {
    // определение типа "прочтённого" элемента (тега)
    int type = reader.next();
    switch (type) {
      case XMLStreamConstants.START_ELEMENT:
      elementName =
         MenuTagName.getElementTagName(reader.getLocalName());
      switch (elementName) {
         case FOOD:
    food = new Food();
    Integer id = Integer.parseInt(
        reader.getAttributeValue( null, "id"));
    food.setId(id);
    break;
      }
    break;
```

StAX. Пример анализа

```
  case XMLStreamConstants.CHARACTERS:
String text = reader.getText().trim();
if (text.isEmpty()) {
    break;
}
switch (elementName) {
case NAME:
    food.setName(text);
    break;
case PRICE:
    food.setPrice(text);
    break;
case DESCRIPTION:
    food.setDescription(text);
    break;
case CALORIES:
    Integer calories = Integer.parseInt(text);
    food.setCalories(calories);
    break;
}
break;
```

StAX. Пример анализа

```
        case XMLStreamConstants.END_ELEMENT:
      elementName =
           MenuTagName.getElementTagName(reader.getLocalName());
      switch (elementName) {
           case FOOD:
           menu.add(food);
      }
         }
      }
     return menu;
   }
}
```

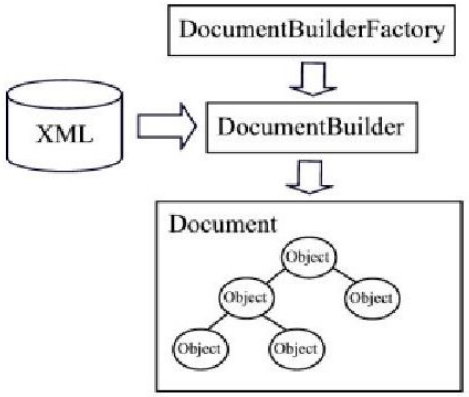
StAX. Пример анализа, запись XML

```
…
XMLOutputFactory factory = XMLOutputFactory. newInstance();

XMLStreamWriter writer = factory.createXMLStreamWriter(
                        new FileWriter("output2.xml"));

writer.writeStartDocument();
writer.writeStartElement("document");
writer.writeStartElement("data");
writer.writeAttribute("name", "value");
writer.writeCharacters("content");
writer.writeEndElement();
writer.writeEndElement();
writer.writeEndDocument();

writer.flush();
writer.close();
…
```

DOM

DOM. Введение

- DOM фундаментально отличается от SAX.
- DOM представляет собой стандарт, а модель DOM не привязана к Java.
- Существуют частные реализации DOM для JavaScript, Java, CORBA и др.



DOM Parsing

DOM. Levels

DOM организован в виде уровней (levels), а не версий. Спецификации DOM можно найти на странице

http://www.w3.org/TR/#tr_DOM

Completed Work

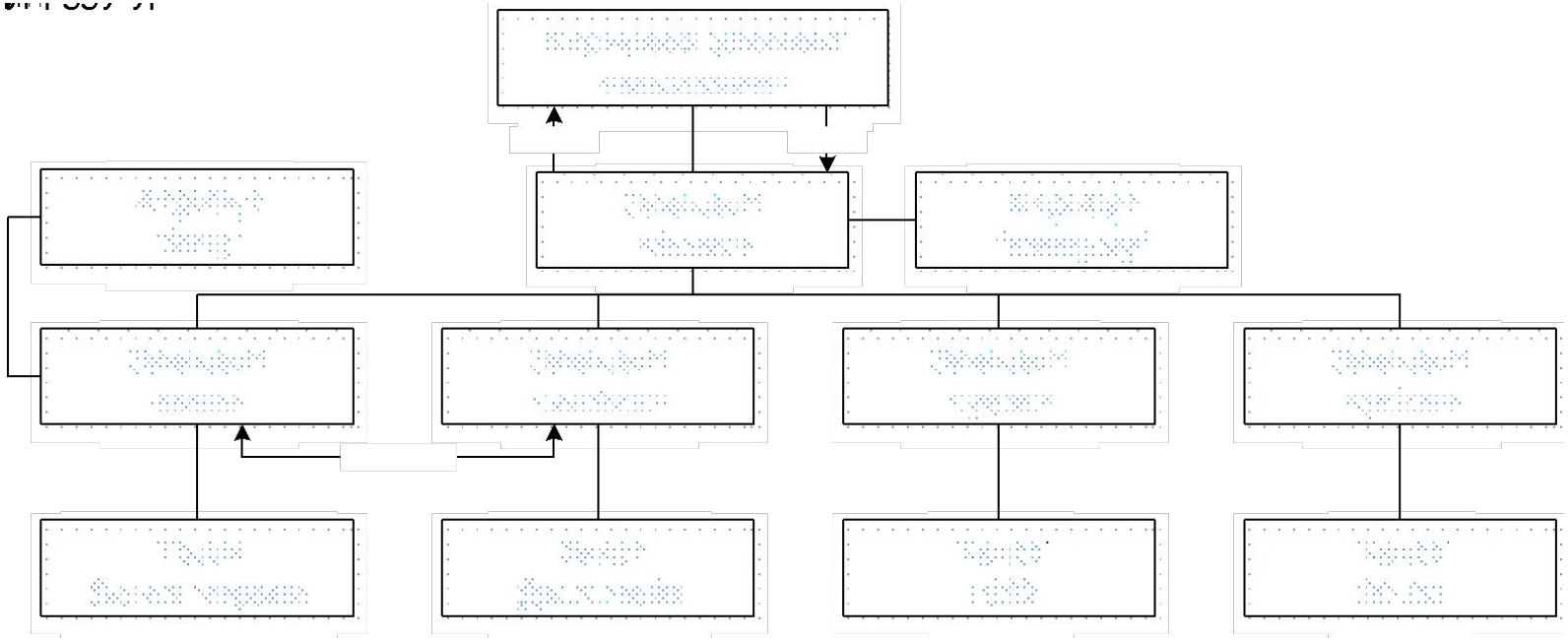
2008-12-22	Element Traversal Specification	Recommendation
2004-04-07	Document Object Model (DOM) Level 3 Load and Save Specification	Recommendation
2004-04-07	Document Object Model (DOM) Level 3 Core Specification	Recommendation
2004-01-27	Document Object Model (DOM) Level 3 Validation Specification	Recommendation
2003-01-09	Document Object Model (DOM) Level 2 HTML Specification	Recommendation
2000-11-13	Document Object Model (DOM) Level 2 Core Specification	Recommendation
2000-11-13	Document Object Model (DOM) Level 2 Events Specification	Recommendation
2000-11-13	Document Object Model (DOM) Level 2 Style Specification	Recommendation
2000-11-13	Document Object Model (DOM) Level 2 Traversal and Range Specification	Recommendation
2000-11-13	Document Object Model (DOM) Level 2 Views Specification	Recommendation
1998-10-01	Document Object Model (DOM) Level 1	Recommendation
2004-02-26	Document Object Model (DOM) Requirements	Group Note
2004-02-26	Document Object Model (DOM) Level 3 Views and Formatting Specification	Group Note
2004-02-26	Document Object Model (DOM) Level 3 XPath Specification	Group Note
2002-07-25	Document Object Model (DOM) Level 3 Abstract Schemas Specification	Group Note

Drafts

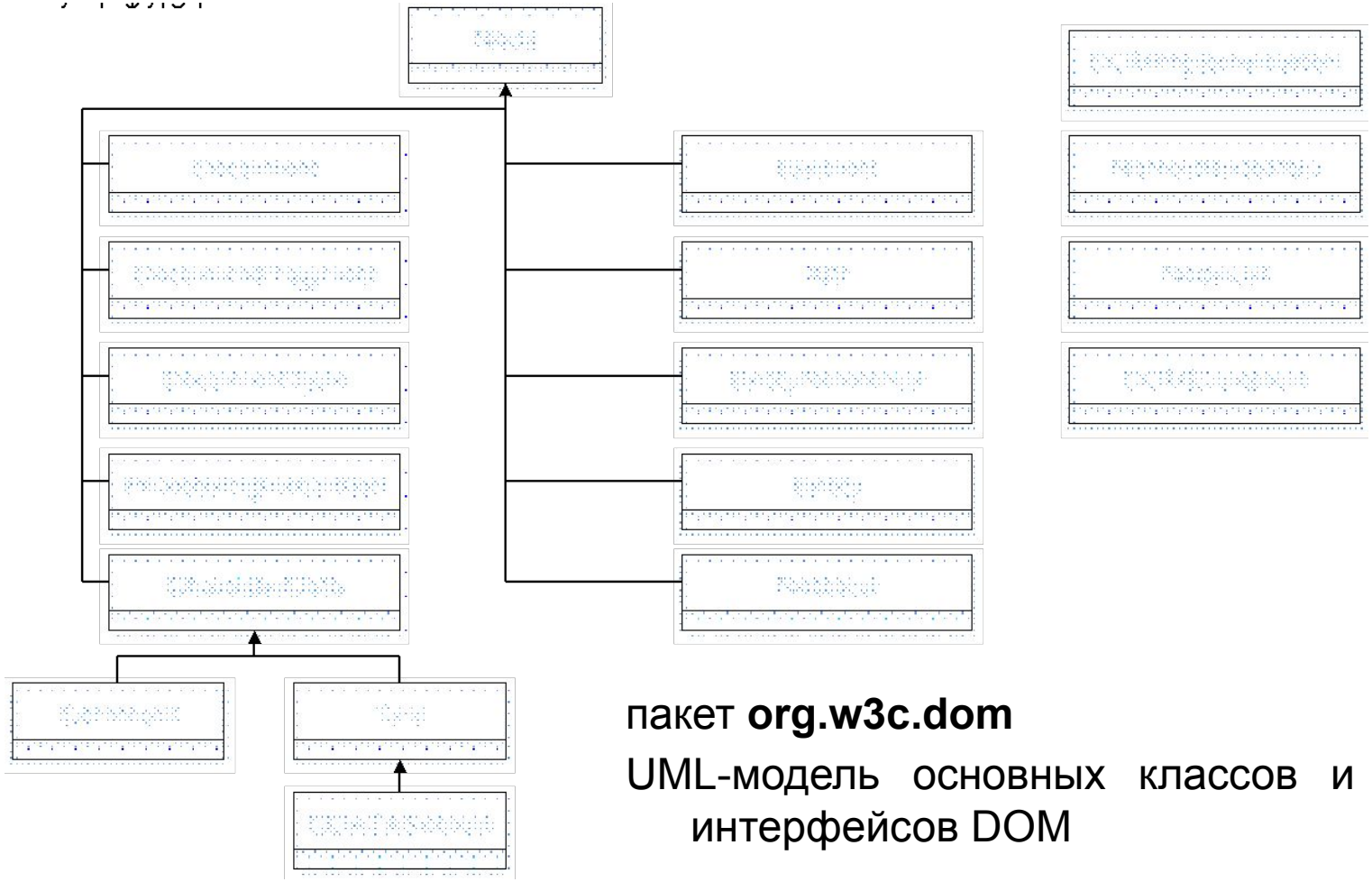
2014-09-25	Document Object Model (DOM) Level 3 Events Specification	Working Draft *Nightly Draft*
2014-07-10	W3C DOM4	Last Call Review ended: 2014-07-31 *Nightly Draft*

DOM. DOM-модель документа

Модель DOM представляет XML-документ как древовидную структуру.



DOM. UML-модель основных классов и интерфейсов



пакет **org.w3c.dom**

UML-модель основных классов и интерфейсов DOM

DOM. Реализации

Существуют различные общепризнанные DOM-анализаторы, которые в настоящий момент можно загрузить с указанных адресов:

Xerces – http://xerces.apache.org/xerces2-j/;

JAXP – входит в JDK.

Существуют также библиотеки, предлагающие свои структуры объектов XML с API для доступа к ним. Наиболее известные:

JDOM – http://www.jdom.org/dist/binary/jdom-1.0.zip.

dom4j – http://www.dom4j.org

DOM. org.w3c.dom.Document

org.w3c.dom.Document

Используется для получения информации о документе и изменения его структуры. Это интерфейс представляет собой корневой элемент XML-документа и содержит методы доступа ко всему содержимому документа.

Метод	Назначение
Element getDocumentElement()	возвращает корневой элемент документа

DOM. org.w3c.dom.Node

org.w3c.dom.Node

Основным объектом DOM является **Node** – некоторый общий элемент дерева. Большинство DOM-объектов унаследовано именно от **Node**. Для представления элементов, атрибутов, сущностей разработаны свои специализации **Node**.

Интерфейс **Node** определяет ряд методов, которые используются для работы с деревом:

Метод	Назначение
short getNodeType()	возвращает тип объекта (элемент, атрибут, текст, **CDATA** и т.д.);
String getNodeValue()	возвращает значение **Node**
Node getParentNode()	возвращает объект, являющийся родителем текущего узла **Node**

DOM. org.w3c.dom.Node

Интерфейс **Node** определяет ряд методов, которые используются для работы с деревом:

Метод	Назначение
NodeList getChildNodes()	возвращает список объектов, являющихся дочерними элементами
Node getFirstChild(), **Node getLastChild()**	возвращает первый и последний дочерние элементы
NamedNodeMap getAttributes()	возвращает список атрибутов данного элемента

У интерфейса **Node** есть несколько важных наследников – **Element**, **Attr**, **Text**. Они используются для работы с конкретными объектами дерева.

DOM. org.w3c.dom.Element

org.w3c.dom.Element

Интерфейс предназначен для работы с содержимым элементов XML-документа. Некоторые методы:

Метод	Назначение
String getTagName(String name)	возвращает имя элемента
boolean hasAttribute()	проверяет наличие атрибутов
String getAttribute(String name)	возвращает значение атрибута по его имени
Attr getAttributeNode(String name)	возвращает атрибут по его имени
void setAttribute(String name, String value)	устанавливает значение атрибута, если необходимо, атрибут создается
void removeAttribute(String name)	удаляет атрибут
NodeList getElementsByTagName(String name)	возвращает список дочерних элементов с определенным именем

DOM. org.w3c.dom.Attr

org.w3c.dom.Attr

Интерфейс служит для работы с атрибутами элемента XML-документа.

Некоторые методы интерфейса **Attr**:

Метод	Назначение
String getName()	возвращает имя атрибута
Element getOwnerElement	возвращает элемент, который содержит этот атрибут
String getValue()	возвращает значение атрибута
void setValue(String value)	устанавливает значение атрибута
boolean isId()	проверяет атрибут на тип ID

DOM. org.w3c.dom.Text

org.w3c.dom.Text

Интерфейс **Text** необходим для работы с текстом, содержащимся в элементе.

Метод	Назначение
String getWholeText()	возвращает текст, содержащийся в элементе
void replaceWholeText(String content)	заменяет строкой **content** весь текст элемента

DOM. Анализ xml-документа с помощью DOM

```
//создание DOM-анализатора (Xerces)

DOMParser parser = new DOMParser();
parser.parse("menu.xml");
Document document = parser.getDocument();

Element root = document.getDocumentElement();

List<Food> menu = new ArrayList<Food>();

NodeList foodNodes = root.getElementsByTagName("food");
Food food = null;
```

DOM. Анализ xml-документа с помощью DOM

```
for (int i = 0; i < foodNodes.getLength(); i++) {
    food = new Food();
    Element foodElement = (Element) foodNodes.item(i);
    food.setId(Integer.parseInt(foodElement.getAttribute("id")));
    food.setName(
     getSingleChild(foodElement, "name").getTextContent().trim());
    food.setDescription(
     getSingleChild(foodElement, "description").getTextContent().trim());
    menu.add(food);
}
```

```
private static Element getSingleChild(Element element, String childName){
    NodeList nlist = element.getElementsByTagName(childName);
    Element child = (Element) nlist.item(0);
    return child;
}
```

DOM. Запись xml-документа с помощью DOM

```
DocumentBuilderFactory factory = DocumentBuilderFactory.newInstance();
DocumentBuilder builder = factory.newDocumentBuilder();
Document document = builder.newDocument();

Element breakfastMenu = document.createElement("breakfast-menu");
Element food = document.createElement("food");
food.setAttribute("id", "234");

Element name = document.createElement("name");
name.setTextContent("Waffles");

food.appendChild(name);
breakfastMenu.appendChild(food);
document.appendChild(breakfastMenu);

TransformerFactory transformerFactory = TransformerFactory.newInstance();
Transformer transformer = transformerFactory.newTransformer();
DOMSource source = new DOMSource(document);
StreamResult result = new StreamResult(new FileWriter("dommenu.xml"));
transformer.transform(source, result);
```

JAXP

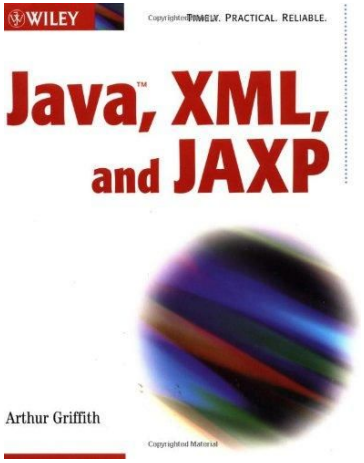
JAXP. Введение

JAXP (JavaAPI for XML Processing) – это API

Он не предоставляет новых способов анализа XML и не обеспечивает функциональности для анализа. Он упрощает работу с некоторыми сложными задачами в DOM и SAX.

JAXP предоставляет способ доступа к API SAX и DOM и работы с результатами анализа документа.

Основная цель JAXP – предоставить независимость от производителя при работе с анализаторами.



JAXP. javax.xml.parsers

JAXP имеет все необходимое для создания как SAX-парсеров, так и DOM-парсеров. В дистрибутив JAXP входит анализатор Sun.

JAXP располагается в пакете **javax.xml.parsers**, в состав которого входят четыре класса:

- **DocumentBuilder** — это DOM-парсер, который создает объект класса org.w3c.dom.Document.
- **DocumentBuilderFactory** — класс, который создает DOM-парсеры.
- **SAXParser** — SAX-парсер, который привязывает обработчик SAX-событий к XML-документу, т.е. обрабатывает XML-документ согласно коду, определенному разработчиком.
- **SAXParserFactory** — класс, который создает SAX-парсеры.

JAXP. javax.xml.parsers

Package javax.xml.parsers

Provides classes allowing the processing of XML documents.

See:

Description

Class Summary	
DocumentBuilder	Defines the API to obtain DOM Document instances from an XML document.
DocumentBuilderFactory	Defines a factory API that enables applications to obtain a parser that produces DOM object trees from XML documents.
SAXParser	Defines the API that wraps an `XMLReader` implementation class.
SAXParserFactory	Defines a factory API that enables applications to configure and obtain a SAX based parser to parse XML documents.

Exception Summary	
ParserConfigurationException	Indicates a serious configuration error.

JAXP. Использование

Чтобы обработать XML-документ с помощью парсеров JAXP, как для SAX, так и для DOM необходимо выполнить 4 действия:

1. Получить ссылку на объект одного из Factory-классов (DocumentBuilderFactory или SAXParserFactory).
2. Настроить необходимые параметры и свойства парсера.
3. Создать парсер.
4. Использовать полученный парсер для обработки XML-документа.

JAXP. Использование

JAXP SAX

```
javax.xml.parsers.SAXParserFactory spf =
            SAXParserFactory.newInstance();
spf.setValidating(false);
javax.xml.parsers.SAXParser sp = spf.newSAXParser();
ConcreteSaxHandler handler = new ConcreteSaxHandler();
sp.parse(new File("menu.xml"), handler);
```

JAXP DOM

```
DocumentBuilderFactory dbf =
        DocumentBuilderFactory.newInstance();
DocumentBuilder db = dbf.newDocumentBuilder();
Document document = db.parse("menu.xml");
```

JAXP. Настройка параметров

Существуют три метода, позволяющих настроить (включить/выключить) некоторые параметры парсера:

- **void setNamespaceAware(boolean awareness)** — если параметр awareness равен true, то будет создан парсер, который будет учитывать пространства имен, если же awareness равен false, тогда пространства имен учитываться не будут.

- **void setValidating(boolean validating)** — если параметр validating равен true, то парсер, перед тем как приступить к обработке XML-документа, сначала проверит его на соответствие его своему DTD.

- **void setFeature(String name, boolean value)** — этот метод позволяет менять некоторые параметры, определяемые производителями парсера, которым вы пользуетесь (т.е., если вы решили воспользоваться парсером от Oracle, JAXP это позволяет). Парсер должен поддерживать спецификацию SAX2.

JAXP. Настройка параметров

Помимо **setValidating** и **setNamespaceAware**, **DocumentBuilderFactory** позволяет также определять следующие параметры:

- **void setCoalescing(boolean value)** — если value равно true, то парсер будет объединять текстовые узлы и CDATA-секции в единые текстовые узлы в DOM-дереве. Иначе CDATA-секции будут вынесены в отдельные узлы.

- **void setExpandEntityReferences(boolean value)** — если установлено (true), то ссылки на сущности будут заменены содержанием этих сущностей (самими сущностями). Если же нет, то эти узлы будут содержать все те же ссылки. Этот метод полезен, если вы не хотите себе головной боли по разыменованию ссылок вручную, если, конечно, они есть.

JAXP. Настройка параметров

- **void setIgnoringComments(boolean value)** — если установлено, то все комментарии, содержащиеся в XML-документе, не появятся в результативном DOM-документе. Если нет, тогда DOM-документ будет содержать узлы с комментариями.

- **void setIgnoringElementContentWhitespace(boolean value)** — если установлено, то пробельные символы (символы табуляции, пробелы и пр.), которые располагаются между элементами XML-документа, будут игнорироваться, и они не будут вынесены в узлы результативного DOM-дерева. Если нет, тогда будут созданы дополнительные текстовые узлы, содержащие эти символы.

В DOM нет метода **setFeature**(), как в SAX2, поэтому установка специальных переменных и параметров здесь не предусмотрена.

Замена анализатора

Смена анализатора фактически означает смену конструктора анализатора, поскольку все экземпляры **SAXParser** и **DocumentBuilder** создаются этими конструкторами.

Для замены реализации интерфейса **SAXParserFactory** установите системное свойство Java

javax.xml.parsers.SAXParserFactory.

Если это свойство не определено, возвращается реализация по умолчанию (анализатор, который указал ваш поставщик).

Аналогичный принцип применим и для используемой вами реализации **DocumentBuilderFactory**. В этом случае запрашивается системное свойство

javax.xml.parsers.DocumentBuilderFactory.

JAXP. TrAX

TrAX (Trasnformation for API) – API для XML преобразований.

TrAX позволяет использовать таблицы стилей XSL для преобразования XML-документов, а также предоставляет возможность преобразования SAX-событий или DOM-документов в XML-файлы и обратно.

javax.xml.transform
javax.xml.transform.dom
javax.xml.transform.sax
javax.xml.transform.stax
javax.xml.transform.stream

```
TransformerFactory tf = TransformerFactory. newInstance();
Templates template = tf.newTemplates(
                new StreamSource("newhello.xsl"));
Transformer transformer = template.newTransformer();
transformer.transform(new StreamSource("hello.xml"),
                new StreamResult("newhello.xml"));
```

JDOM

JDOM. Введение

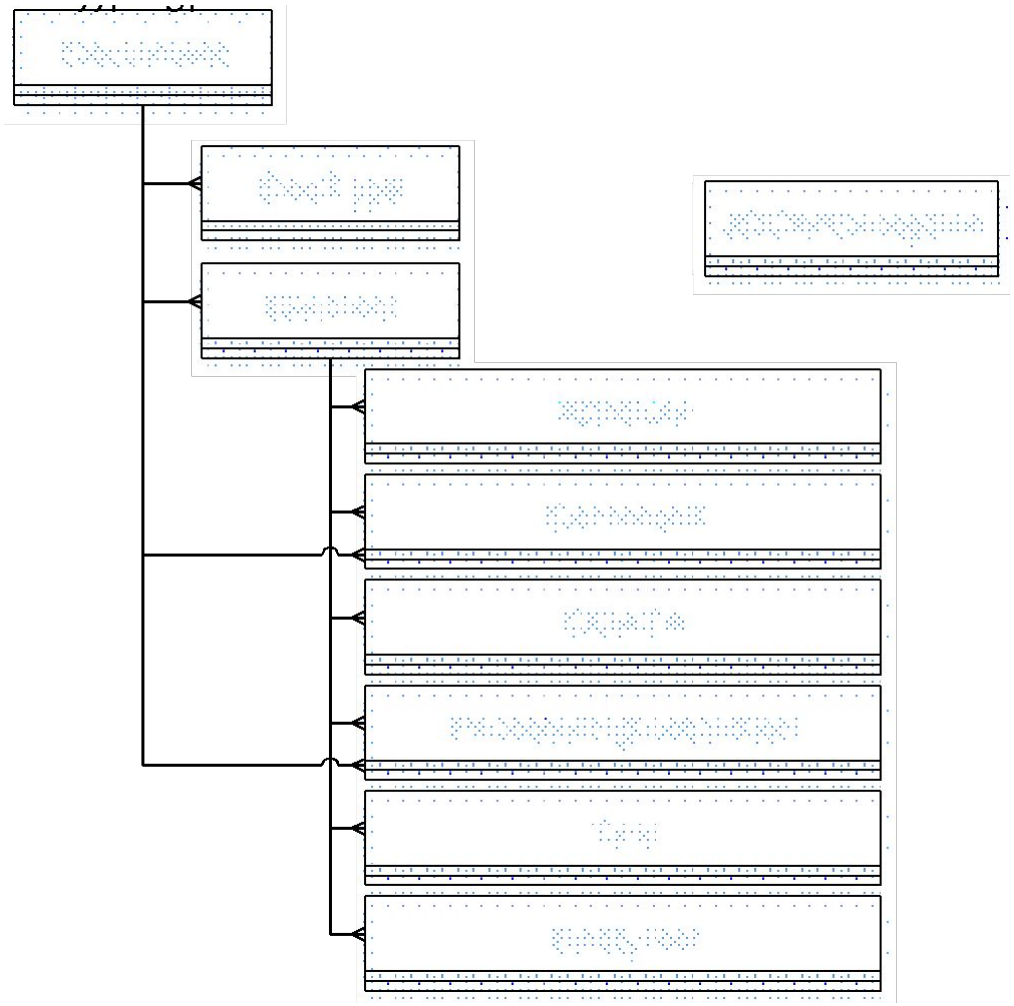
JDOM **не является анализатором**, он был разработан для более удобного, более интуитивного для Java-программист, доступа к объектной модели XML-документа.

JDOM представляет свою модель, отличную от DOM. Для разбора документа JDOM использует либо SAX-, либо DOM-парсеры сторонних производителей.

Реализаций JDOM немного, так как он основан на классах, а не на интерфейсах.

JDOM. UML-модель основных классов JDOM

UML-модель основных классов JDOM



org.jdom2.Content

В корне иерархии наследования стоит класс **Content**, от которого унаследованы остальные классы (**Text**, **Element** и др.).

Основные методы класса **Content**:

Метод	Назначение
Document getDocument()	возвращает объект, в котором содержится этот элемент
Element getParentElement()	возвращает родительский элемент

org.jdom2.Document

Базовый объект, в который загружается после разбора XML-документ. Аналогичен **Document** из Xerces.

Метод	Назначение
Element getRootElement()	возвращает корневой элемент

org.jdom2.Parent

Интерфейс **Parent** реализуют классы **Document** и **Element**. Он содержит методы для работы с дочерними элементами. Интерфейс **Parent** и класс **Content** реализуют ту же функциональность, что и интерфейс **Node** в Xerces.

Некоторые из его методов:

Метод	Назначение
List getContent()	возвращает все дочерние объекты
Content getContent(int index)	возвращает дочерний элемент по его индексу
int getContentSize()	возвращает количество дочерних элементов
Parent getParent()	возвращает родителя этого родителя
int indexOf(Content child)	возвращает индекс дочернего элемента

org.jdom2.Element

Класс **Element** представляет собой элемент XML-документа.

Метод	Назначение
Attribute getAttribute(String name)	возвращает атрибут по его имени
String getAttributeValue(String name)	возвращает значение атрибута по его имени
List getAttributes()	возвращает список всех атрибутов
Element getChild(String name)	возвращает дочерний элемент по имени
List getChildren()	возвращает список всех дочерних элементов
String getChildText(String name)	возвращает текст дочернего элемента
String getName()	возвращает имя элемента
String getText()	возвращает текст, содержащийся в элементе

org.jdom2.Text

Класс **Text** содержит методы для работы с текстом.

Метод	Назначение
String getText()	возвращает значение содержимого в виде строки
String getTextTrim()	возвращает значение содержимого без крайних пробельных символов

org.jdom2.Attribute

Класс **Attribute** представляет собой атрибут элемента XML-документа. В отличие от интерфейса **Attr** из Xerces, у класса **Attribute** расширенная функциональность. Класс **Attribute** имеет методы для возвращения значения определенного типа.

Метод	Назначение
int getAttributeType()	возвращает тип атрибута
тип **get**Тип**Type()**	(**Int**, **Double**, **Boolean**, **Float**, **Long**) возвращает значение определенного типа
String getName()	возвращает имя атрибута
Element getParent()	возвращает родительский элемент

JDOM. Использование

Работа с существующим XML-файлом состоит из следующих этапов:

- Создание экземпляра класса **org.jdom.input.SAXBuilder**, который умеет строить JDOM-дерево из файлов, потоков, URL и т.д.
- Вызов метода **build()** экземпляра SAXBuilder с указанием файла или другого источника.
- Навигация по дереву и манипулирование элементами, если это необходимо.

JDOM. Пример использования

```
SAXBuilder builder = new SAXBuilder();
Document document = builder.build("menu.xml");

Element root = document.getRootElement();
List<Element> menu = root.getChildren();
Iterator<Element> menuIterator = menu.iterator();

while (menuIterator.hasNext()){
    Element foodElement = menuIterator.next();
    System.out.println(foodElement.getChildText("name"));
}
```

JDOM. Изменение XML

JDOM позволяет изменять XML-документ

```
SAXBuilder builder = new SAXBuilder();
Document document = builder.build(filename);
Element root = document.getRootElement();
List c = root.getChildren();
Iterator i = c.iterator();

while (i.hasNext()) {
   Element e = (Element) i.next();
   if (e.getAttributeValue("Ш").equals(login)) {
    e.getChild(element).setText(content);
   }
}

XMLOutputter out = new XMLOutputter();
out.output(document, new FileOutputStream(filename));
```

JDOM. Создание и запись XML

JDOM также позволяет создавать и записывать XML-документы

Для создания документа необходимо создать объект каждого класса (**Element**, **Attribute**, **Document**, **Text** и др.) и присоединить его к объекту, который в дереве XML-документа находится выше.

Element

Для добавления дочерних элементов, текста или атрибутов в элемент XML-документа нужно использовать один из следующих методов:

Метод	Назначение
Element addContent(Content child)	добавляет дочерний элемент
Element addContent(int index, Content child)	добавляет дочерний элемент в определенную позицию
Element addContent(String str)	добавляет текст в содержимое элемента
Element setAttribute(Attribute attribute)	устанавливает значение атрибута
Element setAttribute(String name, String value)	устанавливает атрибут и присваивает ему значение
Element setContent(Content child)	заменяет текущий элемент на элемент, переданный в качестве параметра
Element setContent(int index, Content child)	заменяет дочерний элемент на определенной позиции элементом, переданным как параметр
Element setName(String name)	устанавливает имя элемента
Element setText(String text)	устанавливает текст содержимого элемента

Text

Класс **Text** также имеет методы для добавления текста в элемент XML-документа:

Метод	Назначение
void append(String str)	добавляет текст к уже имеющемуся
void append(Text text)	добавляет текст из другого объекта **Text**, переданного в качестве параметра
Text setText(String str)	устанавливает текст содержимого элемента

Attribute

Методы класса **Attribute** для установки значения, имени и типа атрибута:

Метод	Назначение
Attribute setAttributeType(int type)	устанавливает тип атрибута
Attribute setName(String name)	устанавливает имя атрибута
Attribute setValue(String value)	устанавливает значение атрибута

JDOM. Пример создания нового документа

```
Element root = new Element("breakfast-menu");

Element food = new Element("food");
food.setAttribute("id", "123");

Element name = new Element("name");
name.setText("Waffles");

food.addContent(name);
root.addContent(food);
Document document = new Document(root);

XMLOutputter outputter = new XMLOutputter();
outputter.output(document, new
FileOutputStream("newmenu.xml"));
```

```
<?xml version="1.0"
encoding="UTF-8"?>
<breakfast-menu>
  <food id="123">
    <name>Waffles</name>
  </food>
</breakfast-menu>
```

JAXB

JAXB. Введение

Java Architecture for XML Binding (JAXB) – архитектура связывания данных, обеспечивает связь между XML схемами и Java-представлениями, предоставляя возможность использовать данные представленные в виде XML в приложениях Java.

JAXB предоставляет методы для преобразования XML документов в структуры Java и обратно. Кроме этого, есть возможность генерировать XML схемы из Java объектов.

JAXB. Введение

Используя аннотации JAXB конвертирует объекты в/из XML-файл.

- **Marshalling** – конвертирование java-объектов в XML-файл
- **Unmarshalling** – конвертирование XML в java-объект.

JAXB. Marshalling, пример

```
@XmlRootElement
@XmlAccessorType(XmlAccessType.FIELD)
@XmlType(name = "Food", propOrder = { "name", "price",
"description", "calories" })
public class Food {
    @XmlAttribute(required = true)
    private int id;
    @XmlElement(required = true)
    private String name;
    @XmlElement(required = true)
    private String price;
    @XmlElement(required = true)
    private String description;
    @XmlElement(required = true)
    private int calories;

    public Food(){}
    }
    //  set and get methods
}
```

JAXB. Marshalling, пример

```
JAXBContext context = JAXBContext.newInstance(Food.class);
Marshaller m = context.createMarshaller();

Food food = new Food();
food.setId(123);
food.setName("nnn");
food.setDescription("ddd");
food.setCalories(234);
food.setPrice("333");

m.marshal(food, new FileOutputStream("stud.xml"));
m.marshal(food, System.out);// на консоль
System.out.println("XML-файл создан");
```

JAXB. Unmarshalling, пример

```
File file = new File("stud.xml");
JAXBContext jaxbContext = JAXBContext.newInstance(Food.class);

Unmarshaller jaxbUnmarshaller =
jaxbContext.createUnmarshaller();
Food food = (Food) jaxbUnmarshaller.unmarshal(file);
System.out.println(food.getName());
```

JAXB

Возможно обратное создание на основе XML-схемы классов на языке Java с помощью команды

xjc university.xsd

Больше о JAXB вы можете узнать на https://jaxb.java.net/tutorial/

ВАЛИДАЦИЯ

Валидация

В пакете **javax.xml.validation** для валидации документов используются три класса: **SchemaFactory, Schema** и **Validator**.

Кроме того, этот пакет активно использует интерфейс **javax.xml.transform.Source** для представления документов XML.

```
SchemaFactory factory =
SchemaFactory.newInstance("http://www.w3.org/2001/XMLSchema" );

File schemaLocation = new File("src/resources/notes.xsd" );
Schema schema = factory.newSchema(schemaLocation);
Validator validator = schema.newValidator();
Source source = new StreamSource("src/resources/notes.xml" );
try {
    validator.validate(source);
    System.out.println(" is valid.");
} catch (SAXException ex) {
    System.out.println(" is not valid because " );
    System.out.println(ex.getMessage());
}
```

СОЗДАНИЕ ПРОСТОГО WSDL/SOAP WEB-СЕРВИСА СРЕДСТВАМИ JAVA SE

WSDL/SOAP

- The JDK allows us to both publish and consume a web service using some of its tools. The sample service “Hello world” will be responsible for saying hello to the name that we’ll send it to that service.
- This example also includes creating a client for this service (you can follow the same steps in client to communicate with any service you like).
  - **Creating the Service**
  - **Creating the Client**

WSDL/SOAP: Creating the Service

1. Construct Simple Hello Class

- Suppose you have a simple class that receives a string and return another string

```
public class Hello {
    public String sayHello(String name) {
        return "Hello, " + name;
    }
}
```

WSDL/SOAP : Creating the Service

2. Convert Hello Class to a Web Service

- Simply we can convert this class to be a web service using some annotations
  - @WebService — This identifies the class as being a web service.
  - @SOAPBinding(style=SOAPBinding.Style.RPC) — This specifies the type of the communication, in this case RPC

```
import javax.jws.WebService;
import javax.jws.soap.SOAPBinding;
import static javax.jws.soap.SOAPBinding.Style.RPC;

@WebService
@SOAPBinding(style= RPC)
public class Hello {
    public String sayHello(String name) {
        return "Hello " + name;
    }
}
```

WSDL/SOAP : Creating the Service

3. Publish Hello Service

- To publish this service, we can use the Endpoint class. We will provide the publish method with any URL and an instance of our service class

```
import javax.xml.ws.Endpoint;

public class ServiceStarter {
    public static void main(String[] args) {
        String url = "http://localhost:1212/hello";
        Endpoint.publish(url, new Hello());
        System.out.println("Service started @ " + url);
    }
}
```

WSDL/SOAP : Creating the Service

4. Compile Code

- We can compile our two classes using the simple Javac command:

```
javac -d . *.java
```

5. Start Service

- We can start our service by running ServiceStarter class using the following Java command:

```
java wsserver/ServiceStarter
```

WSDL/SOAP : Creating the Service

6. Check Running Service

- Now the service has been started, you can check your service by seeing its WSDL file by getting the url in setep 3. We can get the Service WSDL file by appending "`?wsdl`" to the URL:

```
http://localhost:1212/hello?wsdl
```

- The result of the WSDL file will look like the following XML file:

WSDL/SOAP : Creating the Service

```
<?xml version="1.0" encoding="UTF-8"?>
<definitions name="HelloService" targetNamespace="http://wsserver/" xmlns:tns="http://wsserver/"
             xmlns="http://schemas.xmlsoap.org/wsdl/"
             xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
             xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <types/>
    <message name="sayHello">
        <part name="arg0" type="xsd:string"/>
    </message>
    <message name="sayHelloResponse">
        <part name="return" type="xsd:string"/>
    </message>
    <portType name="Hello">
        <operation name="sayHello">
            <input message="tns:sayHello"/>
            <output message="tns:sayHelloResponse"/>
        </operation>
    </portType>
    <binding name="HelloPortBinding" type="tns:Hello">
        <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="rpc"/>
        <operation name="sayHello">
            <soap:operation soapAction=""/>
            <input>
                <soap:body use="literal" namespace="http://wsserver/"/>
            </input>
            <output>
                <soap:body use="literal" namespace="http://wsserver/"/>
            </output>
        </operation>
    </binding>
    <service name="HelloService">
        <port name="HelloPort" binding="tns:HelloPortBinding">
            <soap:address location="http://localhost:1212/hello"/>
        </port>
    </service>
</definitions>
```

WSDL/SOAP : Creating the Client

The first thing we should have is an interface of that service class to be able to call its methods using java code. After that we'll write some code to connect to that service. Fortunately there is a tool in JDK called `wsimport` that can do all of that if you just provided it with a valid WSDL URL.

1. Import Service Interface and Service Client Creator Class

- Using `wsimport` tool we will write the following command:

```
wsimport -d . -p wsclient -keep http://localhost:1212/hello?wsdl
```

- The `-p` arg tells the tool to put the generated classes into a specific package. Executing this command will result in generating two classes. The first class, called Hello.java and its interface that contains our method `sayHello`.

WSDL/SOAP : Creating the Client

- The code should be something like this:

```
package com.epam.courses.jf.javase01.mypackage.wsclient;

import javax.jws.WebMethod;
import javax.jws.WebParam;
import javax.jws.WebResult;
import javax.jws.WebService;
import javax.jws.soap.SOAPBinding;

/**
 * This class was generated by the JAX-WS RI.
 * JAX-WS RI 2.1.6 in JDK 6
 * Generated source version: 2.1
 */

@WebService(name = "Hello", targetNamespace = "http://wsserver/")
@SOAPBinding(style = SOAPBinding.Style.RPC)
public interface Hello {
    /**
     * @param arg0
     * @return returns java.lang.String
     */
    @WebMethod
    @WebResult(partName = "return")
    public String sayHello(@WebParam(name = "arg0", partName = "arg0") String arg0);
}
```

WSDL/SOAP : Creating the Client

- The second file would be called HelloService.java, and it will contain the methods that would help us to connect to our service we are only concerned with the no-arg constructor and the `getHelloPort()` method:

```
@WebServiceClient (name = "HelloService", targetNamespace = "http://wsserver/",
                   wsdlLocation = "http://localhost:1212/hello?wsdl")
public class HelloService extends Service {
    private final static URL HELLOSERVICE_WSDL_LOCATION;
    private final static Logger logger = Logger.getLogger(wsclient.HelloService.class.getName());

    //...

    public HelloService() {
        super(HELLOSERVICE_WSDL_LOCATION, new QName("http://wsserver/", "HelloService"));
    }

    /** @return returns Hello */
    @WebEndpoint (name = "HelloPort")
    public Hello getHelloPort() {
        return super.getPort(new QName("http://wsserver/", "HelloPort"), Hello.class);
    }

    /** @param features A list of {@link javax.xml.ws.WebServiceFeature} to configure on the proxy.
     * Supported features not in the <code>features</code> parameter will have their default values.
     * @return returns Hello */
    @WebEndpoint (name = "HelloPort")
    public Hello getHelloPort(WebServiceFeature... features) {
        return super.getPort(new QName("http://wsserver/", "HelloPort"), Hello.class, features);
    }
}
```

WSDL/SOAP : Creating the Client

2. Invoke the Web Service

- We are now ready to write the code responsible for invoking the web service by making a new instance of the `HelloService` class, we are ready to get Hello interface by calling the method `getHelloPort()` from the `HelloService` instance. After that we can call the method and get the response as a simple Java method:

```
public class HelloClient {
    public static void main(String[] args) {
        HelloService service = new HelloService();
        Hello hello = service.getHelloPort();
        String text = hello.sayHello( "Henry");
        System.out.println(text);
    }
}
```

3. Compile Classes and Run

```
javac -d . *.java
java wsclient/HelloClient
```

JAVASCRIPT OBJECT NOTATION (JSON)

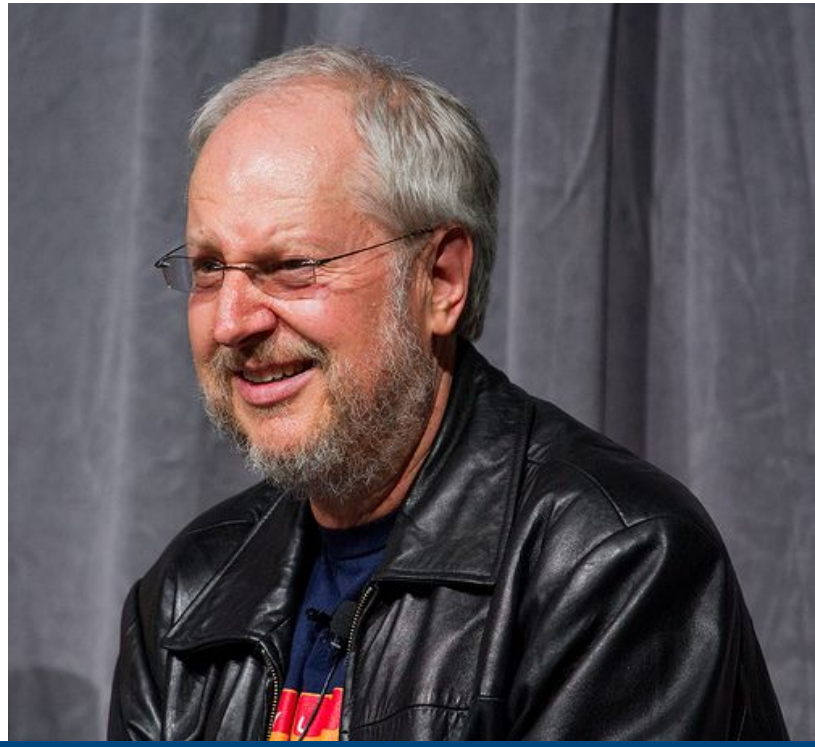
JSON

1. XML изначально создавался как **метаязык разметки документов**, позволяя использовать унифицированный код парсера и валидатора документов.
2. Однако, будучи первым стандартом такого рода, да еще и пришедшимся на период бурного внедрения цифровых корпоративных информационных систем, XML послужил основой для бесчисленного множества стандартов сериализации данных и протоколов взаимодействия, т.е. хранения и передачи структурированных данных.
3. Тогда как **создавался он прежде всего для разметки документов**.
4. Будучи разрабатываемым комитетами, стандарт XML оказался дополнен множеством расширений, позволяющих, в частности, избегать конфликтов имен и выполнять сложные запросы в XML-документах.
5. Поскольку получающееся нагромождение тэгов оказывалось совершенно нечитаемым никаким человеком, был разработан и широко реализован стандарт XML Schema, позволяющий на том же XML абсолютно строго описать допустимое содержимое каждого документа с целью последующей автоматической проверки.

JSON

- Тем временем, все больше разработчиков под влиянием зарождающихся интерактивных web-технологий стало знакомиться с языком JavaScript, и они начали осознавать, что для представления структурированных объектов в текстовом виде совершенно не обязательно изучать много сотен страниц XML-спецификаций.

- **Дуглас Крокфорд** предложил стандартизовать подмножество JavaScript для сериализации объектов (но не разметки документов!) безотносительно к языку, идея была поддержана сообществом. В настоящее время JSON является одним из двух (вместе с XML) языков, поддерживаемых всеми сколько-либо популярными технологиями программирования.



JSON - Проблемы XML`а как базы для интеграции систем

- **Недостаточная лаконичность**
  - При закрытии тега нужно указывать его – зачастую довольно длинное, квалифицированное имя.
  - Альтернативой выступают значительно более лаконичные атрибуты тегов, но их использование резко ограничивают
    - Атрибуты, согласно большинству стайлгайдов, рекомендуется использовались практически исключительно для метаинформации (например, id), так что основной контент обычно формировался тегами
    - Неймспейсы на практике практически невозможно применять к атрибутам
- **Ориентация на документы, а не на сообщения**
- **Символьный формат**

JSON - Проблемы XML`а как базы для интеграции систем

```
{
  "firstName": "Иван",
  "lastName": "Иванов",
  "address": {
    "streetAddress": "Московское ш., 101, кв.101",
    "city": "Ленинград",
    "postalCode": 101101
  },
  "phoneNumbers": [
    "812 123-1234",
    "916 123-4567"
  ]
}
```

JSON Schema

1. Массово начав использовать JSON для представления данных, разработчики столкнулись с необходимостью вручную проверять содержимое документов, каждый раз на каждом языке переизобретая логику валидации.
2. Людей, знакомых с **XML Schema**, это не могло не раздражать. И постепенно аналогичный стандарт JSON Schema был-таки сформирован и располагается по адресу http://json-schema.org
3. http://json-schema.org/implementations.html - список open-source библиотек для различных ЯП
4. Для Java
   - json-schema-validator
   - json-schema
   - json-schema-validator

JSON Schema

Пример простой схемы, задающей словарь 2D или 3D геометрических точек в пространстве
(-1, 1) x (-1, 1) x (-1, 1)
с ключами, состоящими из цифр

```
{
 "type": "object",
 "patternProperties": {
  "^[0-9]+$": {
   "type": "object",
   "properties": {
    "value": {
     "type": "number",
     "minimum": 0
    },
    "x": {"$ref": "#/definitions/point_coord"},
    "y": {"$ref": "#/definitions/point_coord"},
    "z": {"$ref": "#/definitions/point_coord"}
   },
   "required": ["value", "x", "y"]
  }
 },
 "additionalProperties": false,
 "definitions": {
  "point_coord": {
   "type": "number",
   "minimum": -1,
   "maximum": 1
  }
 }
}
```

СПАСИБО ЗА ВНИМАНИЕ!

ВОПРОСЫ?

Java.SE.13

Integration Data Formats

Author: **Olga Smolyakova , PhD**
Oracle Certified Java 6 Programmer
Olga_Smolyakova@epam.com