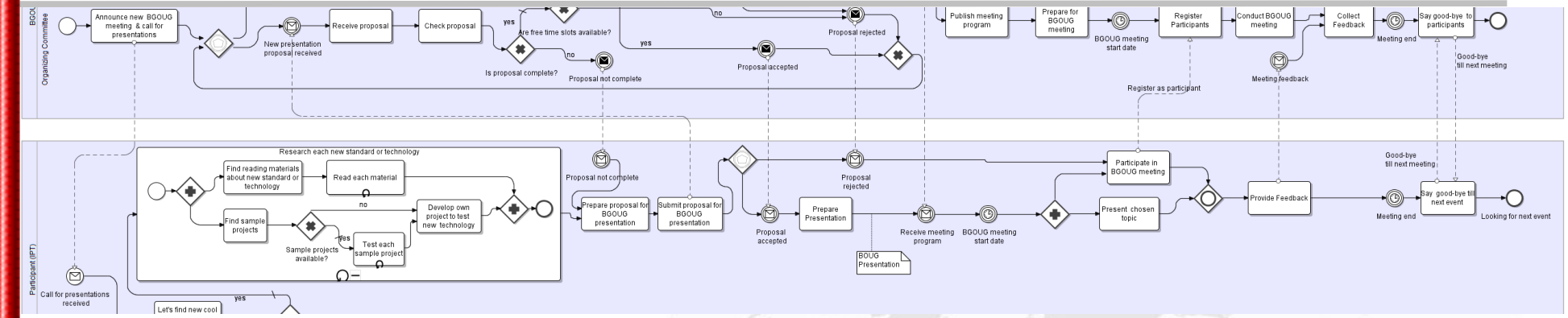
Trayan Iliev
IPT – Intellectual Products & Technologies Ltd.

Java™ API for XML Processing (JAXP). Streaming API for XML (StAX). Java Architecture for XML Binding (JAXB)



Траян Илиев

IPT – Intellectual Products & Technologies
e-mail: tiliev@iproduct.org
web: http://www.iproduct.org

Trayan Iliev
IPT – Intellectual Products & Technologies Ltd.

Съдържание

Програмни интерфейси за обработка на XML

- Simple API for XML (SAX) – събитийно ориентиран интерфейс, при който документът се чете последователно и неговото съдържание се рапортува като извиквания на отделни **callback** методи на обработващия обект (**handler**)
- Document Object Model (DOM) – зарежда и манипулира в паметта целия документ под формата на дърво
- Pull Parsing: Streaming API for XML (StAX) – третира документа като серия от възли, които се обхождат последователно с помощта на шаблона **Iterator**
- Свързване на данни (Data Binding): Java Architecture for XML Binding (JAXB) – дава възможност да се автоматизира процесът на запазване/извличане на данните от/в XML формат (marshalling/unmarshalling)

Java™ API for XML Processing (JAXP)

- Simple API for XML (SAX)
  - Пакет: `org.xml.sax`
  - Factory class: SAXParserFactory
  - Parser interfaces: SAXParser обвива SAXReader
- Общи интерфейси към различните имплементации на SAX и DOM парсъри – пакет `javax.xml.parsers`

Java API for XML Processing (JAXP) (2)

- Document Object Model (DOM)
  - Пакет: `org.w3c.dom`
  - Factory class: DocumentBuilderFactory
  - Document model builder class: DocumentBuilder
- Extensible Stylesheet Language Transformations (XSLT)
  - Пакет: `javax.xml.transform`
  - Factory class: TransformerFactory
  - XML transformer class: Transformer

Simple API for XML (SAX)

- SAX пакети:
  - **org.xml.sax** - SAX интерфейси
  - **org.xml.sax.ext** - SAX2 разширения
  - **org.xml.sax.helpers** – класове улесняващи имплементацията на SAX парсъри (клас DefaultHandler)
  - **javax.xml.parsers** – клас SAXParserFactory, който връща SAXParser, Exceptions

Simple API for XML (SAX) (2)

- SAX основни класове и интерфейси:
  - SAXParserFactory
  - SAXParser
  - SAXReader
  - DefaultHandler
  - ContentHandler
  - ErrorHandler
  - ErrorHandler
  - EntityResolver
  - EntityResolver2
  - LexicalHandler
  - Attributes
  - Attributes2
  - Locator
  - Locator2
  - DeclHandler

Document Object Model (DOM)

- DOM пакети:
  - **org.w3c.dom** - DOM програмни интерфейси за достъп и манипулиране на данни на XML елементите, дефиниран от W3C.
  - **javax.xml.parsers** – общи класове за достъп до различните DOM имплеметации: **DocumentBuilderFactory** и **DocumentBuilder** конкретната имплементация се определя от системното свойство (property) **javax.xml.parsers.DocumentBuilderFactory**, **Exceptions**

Document Object Model (DOM) (2)

- DOM основни класове и интерфейси:
  - DocumentBuilderFactory
  - DocumentBuilder
  - Attr
  - CDATASection
  - CharacterData
  - Comment
  - Document
  - DocumentFragment
  - DocumentType
  - DOMConfiguration
  - DOMError
  - DOMErrorHandler
  - DOMLocator
  - DOMStringList
  - Element
  - Entity

Document Object Model (DOM) (3)

- DOM основни класове и интерфейси:
  - EntityReference
  - NamedNodeMap
  - NameList
  - Node
  - NodeList
  - Notation
  - ProcessingInstruction
  - Text
  - TypeInfo
  - UserDataHandler

Extensible Stylesheet Language Transformations (XSLT)

- XSLT пакети:
  - **javax.xml.transform** - TransformerFactory и Transformer класове – transform(source, result)
  - **javax.xml.transform.dom** – DOM sources и results
  - **javax.xml.transform.sax** – SAX sources и results
  - **javax.xml.transform.stax** – Streaming API for XML
  - **javax.xml.transform.stream** – I/O stream sources и results

Extensible Stylesheet Language Transformations (XSLT) (2)

- DOM основни класове и интерфейси:
  - TransformerFactory
  - Transformer
  - ErrorListener
  - Result
  - Source
  - SourceLocator
  - Templates
  - URIResolver
  - SAXResult
  - SAXSource
  - DOMResult
  - DOMSource
  - StAXResult
  - StAXSource
  - StreamResult
  - StreamSource

Упражнения

...

Програмни интерфейси за обработка на XML

- Simple API for XML (SAX) – събитийно ориентиран интерфейс, при който документът се чете последователно и неговото съдържание се рапортува като извиквания на отделни **callback** методи на обработващия обект (**handler**)
- Document Object Model (DOM) – зарежда и манипулира в паметта целия документ под формата на дърво
- Pull Parsing: Streaming API for XML (StAX) – третира документа като серия от възли, които се обхождат последователно с помощта на шаблона **Iterator**
- Свързване на данни (Data Binding): Java Architecture for XML Binding (JAXB) – дава възможност да се автоматизира процесът на запазване/извличане на данните от/в XML формат (marshalling/unmarshalling)

Streaming API for XML (StAX)

- Java базиран, събитийно-ориентиран, Pull Parsing API
- Позволява както четене така и писане на XML документи (двупосочно парсване)
- Бързина, изисква малко памет, сравнително лесно програмиране (шаблон “Итератор”)
- При Pull Parsing няколко различни документа могат да бъдат обработвани едновременно с една единствена нишка
- Две API, които се допълват взаимно:
  - Cursor API – бързо е и използва минимално количество памет, данните се получават като примитивни типове
  - Iterator API – данните са под формата на Event обекти и са по-лесни за обработка и предаване/запазване

Streaming API for XML (StAX)

StAX пакети:

- **javax.xml.stream** - StAX програмни интерфейси и фабрики за обекти позволяващи четене и писане на XML чрез Cursor API и Iterator API
  - Основни класове: XMLEventFactory, XMLInputFactory, XMLOutputFactory
  - Основни интерфейси: XMLStreamReader, XMLStreamWriter, XMLEventReader, XMLEventWriter
- **javax.xml.stream.events** – StAX програмни интерфейси моделиращи основните типове събития използвани от Iterator API
- **javax.xml.stream.util** – допълнителни помощни класове за StAX

Streaming API for XML - типове събития

- StartDocument
- StartElement
- EndElement
- StartElement.
- Characters
- EntityReference

- ProcessingInstruction
- Comment
- EndDocument
- DTD
- Attribute
- Namespace

StAX пример 1 (1)

```
public class StaxParsingCursor {
public static void main(String[] args) {
   XMLInputFactory inputFactory = null;
   try {
      inputFactory = XMLInputFactory.newFactory();
      inputFactory.setProperty(XMLInputFactory.IS_COALESCING, false);
      inputFactory.setProperty(XMLInputFactory.IS_SUPPORTING_EXTERNAL_ENTITIES,
         true);
      inputFactory.setProperty(XMLInputFactory.IS_REPLACING_ENTITY_REFERENCES,
         true);
   } catch (Exception e) {e.printStackTrace();}
   System.out.println("INPUT FACTORY: " + inputFactory);
   try {
      XMLStreamReader streamReader = inputFactory.createXMLStreamReader(
          new FileInputStream("demo.xml"));
```

StAX пример 1 (2)

```
if (streamReader.getEventType() == XMLStreamReader.START_DOCUMENT) {
   System.out.println("<?xml version=\"" + streamReader.getVersion() + "\""
    + " encoding=\"" + streamReader.getCharacterEncodingScheme() + "\"?>");
}
while (streamReader.hasNext()) {
   int parsingEventType = streamReader.next();
   if (streamReader.isStartElement()) {
      System.out.print("<" + streamReader.getName());
      int count = streamReader.getAttributeCount();
      for (int i = 0; i < count; i++) {
         System.out.print(" " + streamReader.getAttributeName(i) + "=\""
                          + streamReader.getAttributeValue(i) + "\"");
      }
      System.out.print(">");
   } else if (streamReader.isEndElement()) {
      System.out.print("</" + streamReader.getName() + ">");
```

StAX пример 1 (3)

```
            } else if (streamReader.hasText()) {
                    System.out.print(streamReader.getText());
            }
        }
    } catch (XMLStreamException e) {
        System.out.println(e.getMessage());
            if (e.getNestedException() != null) {
                e.getNestedException().printStackTrace();
             }
        } catch (Exception ex) {
        ex.printStackTrace();
    }
}
```

StAX пример 2 (1)

```
public class XHTMLGenarator{
    public static void main(String[] args) throws Exception {
        XMLOutputFactory xof = XMLOutputFactory.newFactory();
        FileOutputStream fos = new FileOutputStream("output.xhtml");
        XMLStreamWriter xsw = xof.createXMLStreamWriter(fos,"utf-8");
        xsw.writeStartDocument("utf-8", "1.0");
        xsw.setPrefix("html", "http://www.w3.org/1999/xhtml");
        xsw.writeComment("XHTML document format");
        xsw.writeStartElement("http://www.w3.org/1999/xhtml", "html");
        xsw.writeNamespace("html", "http://www.w3.org/1999/xhtml");
        xsw.writeStartElement("http://www.w3.org/1999/xhtml", "head");
        xsw.writeStartElement("http://www.w3.org/1999/xhtml", "title");
        xsw.writeCharacters("RepRap 3D Printer");
        xsw.writeEndElement(); xsw.writeEndElement();
```

StAX пример 2 (2)

```
            xsw.writeStartElement("http://www.w3.org/1999/xhtml", "body");
            xsw.writeStartElement("http://www.w3.org/1999/xhtml", "p");
            xsw.writeCharacters("You can get more information about RepRap
                at ");
            xsw.writeStartElement("http://www.w3.org/1999/xhtml", "a");
            xsw.writeAttribute("href", "http://reprap.org/wiki/RepRap");
            xsw.writeCharacters("this site");
            xsw.writeEndElement(); xsw.writeEndElement();
            xsw.writeEndElement(); xsw.writeEndDocument();
            xsw.writeEndElement();
            xsw.close();
        }
}
```

StAX пример 3 – Event API

```
public static void main(String[] args) throws Exception {
    XMLInputFactory factory = XMLInputFactory.newInstance();
    System.out.println("Created factory: " + factory);
    XMLEventReader reader = factory.createXMLEventReader(
        new FileInputStream("myDoc.xml"));
    while (reader.hasNext()) {
        XMLEvent e = reader.nextEvent();
        System.out.println(getEventTypeString(e.getEventType())+ ":" +
e);
    }
}
```

Програмни интерфейси за обработка на XML

- Simple API for XML (SAX) – събитийно ориентиран интерфейс, при който документът се чете последователно и неговото съдържание се рапортува като извиквания на отделни **callback** методи на обработващия обект (**handler**)
- Document Object Model (DOM) – зарежда и манипулира в паметта целия документ под формата на дърво
- Pull Parsing: Streaming API for XML (StAX) – третира документа като серия от възли, **Iterator**
- Свързване на данни (Data Binding): **Java Architecture for XML Binding (JAXB)** – дава възможност да се автоматизира процесът на запазване/извличане на данните от/в XML формат (marshalling/unmarshalling)

Java™ Architecture for XML Binding (JAXB)

- Дава възможност да се автоматизира процесът на запазване/извличане на данните от/в Java обекти (POJO) към/от XML формат (marshalling/unmarshalling)
- Използва XML Schema като формат за описание на структурата на данните и от него генерира Java класове с JAXB анотации (xjc команда)
- JAXB анотираните класове се използват за автоматично създаване на обекти по време на изпълнение – чрез unmarshaling (десериализация) от XML файл и след това за marshaling (сериализация) обратно от java обектите в XML
- Поддържа се и обратната посока - от JAXB анотираните класове се генерира XML Schema за валидация (schemagen)

Основни компоненти на JAXB

- XML Schema компилатор: xjc.exe (в bin директорията на Java™ EE сървъра) – генерира JAXB анотирани Java™ класове и пакети:

```
C:\NotesXMLDB>xjc -xmlschema note.xsd -p notification.jaxb -d src
```

- XML Schema генератор: schemagen.exe (в bin директорията на Java™ EE сървъра) – генерира JAXB анотирани Java™ класове и пакети:

```
C:\NotesXMLDB>schemagen -d generated notification.jaxb.Note
notification.jaxb.Notes notification.jaxb.ObjectFactory notification\
jaxb\package-info.java
```

- Binding Runtime Framework – осъществява unmarshaling /marshaling на java обектите в XML

Пакети на JAXB

- javax.xml.bind – основни класове реализиращи свързване на java обекти и xml елементи за клиентските приложения, включително marshaling (сериализация) / unmarshaling (десериализация) и валидация
- javax.xml.bind.annotation – основни анотации позволяващи свързване (mapping) между java обекти и xml елементи
- javax.xml.bind.annotation.adapters – дефинира адаптерен клас XmlAdapter<ValueType,BoundType>, който позволява да кажем на JAXB как да сериализира / десериализира определени класове с помощта на анотацията @XmlJavaTypeAdapter
- javax.xml.bind.attachment – поддръжка на двоично кодирани прикачени части (attachments)
- javax.xml.bind.helpers – помощни класове (собствен JAXB Provider)

Основни анотации на JAXB (1)

- @XmlAccessorOrder – задава реда на сериализация на JAXB свойствата Java клас или пакет
- @XmlAccessorType – контролира дали се сериализират полета или свойства на Java класа
- @XmlAnyAttribute – обозначава Java поле или свойство от тип java.util.Map, което ще получи всички атрибути, за които не е специфицирана @XmlAttribute анотация
- @XmlAnyElement – обозначава Java поле или свойство от тип списък, което ще получи всички елементи, за които не са специфицирани @XmlElement или @XmlElementRef анотации (съответства на XML схема тип xs:any)
- @XmlAttribute – задава съответствие на поле или свойство на Java класа с атрибут в XML файла

Основни анотации на JAXB (2)

- @XmlElement – задава съответствие на поле или свойство на Java класа с елемент в XML файла (вкл. Namespace, подразб. ст.)
- @XmlElementDecl – свързва XML схема декларация на елемент и метод на ObjectFactory клас, който създава елемента (от тип JAXBElement)
- @XmlElementRef – позволява динамично асоцииране на имена на XML елементи с JavaBean свойства
- @XmlElementRefs – групира @XmlElementRef анотации
- @XmlElements – групира @XmlElement анотации
- @XmlElementWrapper – генерира „обвиващ“ елемент около свойство от тип колекция от стойности (списък, множество, ...)
- @XmlEnum – задава enum клас включващ @XmlEnumValue стойности и задава базовия тип на XML схема типа

Основни анотации на JAXB (3)

- @XmlInlineBinaryData – изключва възможността за оптимизирано двоично кодиране (XOP) на съответното свойство и специфицира че двоичните данни ще бъдат кодирани с base64 encoding
- @XmlList – позволява множество стойности да бъдат специфицирани в общ XML елемент, разделени с интервали
- @XmlMimeType – задава MIME тип на base64-encoded двоични данни (напр. Image или Source)
- @XmlMixed – задава, че съответното свойство съответства на XML елемент с Mixed Content модел
- @XmlNs – специфицира съответствие на пространство от имена и префикс (например на ниво Java пакет)
- @XmlRegistry – анотира клас като фабрика за обекти (ObjectFactory)

Основни анотации на JAXB (4)

- **@XmlRootElement** – често използвана анотация, която декларира коренния елемент на XML документ; може да се приложи към Java™ клас или enumeration
- @XmlSchema – анотира Java™ пакет, като дефинира параметрите на схемата, която пакетът реализира (targetNamespace, elementFormDefault, attributeFormDefault, допълнителни пространства от имена и префикси)
- @XmlSchemaType - свързва клас с прост, вграден XML схема тип
- @XmlTransient -изключва даденото поле от JAXB сериализацията
- @XmlType – дефинира и настройва детайлите на свързването на Java™ клас или enumeration с XML схема тип
- @XmlValue – свързва клас към XML схема сложен тип с просто съдържание (complex type with a simpleContent) или прост тип

JAXB пример: XML схема

```
<complexType>
   <complexContent>
      <restriction base="xs:anyType">
         <sequence>
            <element name="id" type="xs:long"/>
            <element name="to" type="xs:string"/>
            <element name="from" type="xs:string"/>
            <element name="heading" type="xs:string"/>
            <element name="body" type="xs:string"/>
         </sequence>
       </restriction>
    </complexContent>
</complexType>
```

JAXB пример: JAXB клас

```
@XmlAccessorType(XmlAccessType.FIELD)
@XmlType(name = "", propOrder = {"id", "to", "from", "heading", "body"})
@XmlRootElement(name = "note")
public class Note {
    protected long id;
    @XmlElement(required = true)
    protected String to;
    @XmlElement(required = true)
    protected String from;
    @XmlElement(required = true)
    protected String heading;
    @XmlElement(required = true)
    protected String body;
...
}
```

Упражнения

...

Референции

- J2EE 1.4 Tutorial – http://java.sun.com/j2ee/1.4/docs/tutorial/doc/index.html
- Java EE 5 Tutorial – http://java.sun.com/javaee/5/docs/tutorial/doc/
- W3Schools – уроци за XML технологии: http://w3schools.com/

Благодаря Ви за Вниманието!

Въпроси?