

Semi-structured Data

6 - Simple API for XML (SAX) Methods Overview

ContentHandler Methods

`public void startDocument() throws SAXException`

is called just once at the **beginning of parsing**

`public void endDocument() throws SAXException`

is called just once, and is the **last method called by the parser**

ContentHandler Methods

```
public void processingInstruction(String target, String data)  
    throws SAXException
```

is called once for **each processing instruction** `<?target data?>`

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

target: `xml-stylesheet`

data: `href="course.css" type="text/css"`

ContentHandler Methods

```
public void startElement(String namespaceURI,  
                        String localName,  
                        String qualifiedName,  
                        Attributes atts) throws SAXException
```

is called once at the **beginning of every element**

- If the parser **is namespace-aware** (later)
 - namespaceURI holds the prefix (prefix:localname)
 - localName holds the element name (without a prefix)
 - qualifiedName is empty
- If the parser **is not namespace-aware**
 - namespaceURI, localName are empty
 - qualifiedName holds the element name (possibly with a prefix)

Attributes

- When SAX calls startElement, it passes in a parameter of **type Attributes**
- Attributes is an interface that defines some useful methods:
 - `getLength()` - number of attributes
 - `getLocalName(index)` - attribute's local name
 - `getQName(index)` - attribute's qualified name
 - `getValue(index)` - attribute's value
 - `getType(index)` - attribute's type (CDATA, NMTOKEN, etc.)

ATTENTION: If local name is empty, then qualified name hold's the attribute's name

ATTENTION: SAX does not guarantee the order of the attributes

ContentHandler Methods

```
public void endElement(String namespaceURI,  
                      String localName,  
                      String qualifiedName) throws SAXException
```

is called once at the **end of every element**

- If the parser **is namespace-aware** (later)
 - namespaceURI holds the prefix (prefix:localname)
 - localName holds the element name (without a prefix)
 - qualifiedName is empty
- If the parser **is not namespace-aware**
 - namespaceURI, localName are empty
 - qualifiedName holds the element name (possibly with a prefix)

ContentHandler Methods

```
public void characters(char[] ch,  
                      int start,  
                      int length) throws SAXException
```

is called to pass in **character data**

```
<?xml version="1.0"?>  
<course>  
  <acronym>SSD</acronym>Semi-structured Data  
</course>
```

```
ch: <?xml version="1.0"?>  
    <course>  
        <acronym>SSD</acronym>Semi-structured Data  
    </course>
```

start: 43

SSD

length: 3

ContentHandler Methods

```
public void characters(char[] ch,  
                      int start,  
                      int length) throws SAXException
```

is called to pass in **character data**

```
<?xml version="1.0"?>  
<course>  
  <acronym>SSD</acronym>Semi-structured Data  
</course>
```

```
ch: <?xml version="1.0"?>  
<course>  
  <acronym>SSD</acronym>Semi-structured Data  
</course>
```

start: 56

Semi-structured Data

length: 20

ContentHandler Methods

```
public void characters(char[] ch,  
                      int start,  
                      int length) throws SAXException
```

is called to pass in **character data**

The **string constructor** can be used to extract the relevant characters

```
new String(ch, start, length)
```