

# XML Documents



# eXtensible Markup Language

- A textual document format used all over the web is **XML**
  - Used to represent ***hierarchically organized*** data for “ease of use” both by humans and computers
  - Had its origins in SGML (Standard Generalized Markup Language) from the 1980s
  - Became a standard in the late 1990s

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
  <author>Cay Horstmann</author>
  <title>
    Java for Everyone: Late Objects
  </title>
  ...
</book>
```

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="078-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
  <author>Cay Horstmann</author>
  <title>
    Java for Everyone: Late Objects
  </title>
  ...
</book>
```

An **XML declaration** is the first line of well-formed XML file.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
    <author>Cay Horstmann</author>
    <title>
        Java for Everyone: Late Objects
    </title>
    ...
</book>
```

This is not part of the actual document/file!

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
    <author>Cay Horstmann</author>
    <title>
        Java for Everyone: Late Objects
    </title>
    ...
</book>
```

This is an  
author  
**element**.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
    <author>Cay Horstmann</author>
    <title>
        Java for Everyone: Late Objects
    </title>
    ...
</book>
```

This is a  
***start tag***  
for an author  
element.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
    <author>Cay Horstmann</author>
    <title>
        Java for Everyone: Late Objects
    </title>
    ...
</book>
```

This is an  
**end tag**  
for that author  
element.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-1"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
    <author>Cay Horstmann</author>
    <title>
        Java for Everyone: Late Objects
    </title>
    ...
</book>
```

The **content** for this author element is everything between its start and end tags.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20, 2011">
    <author></author>
    <title>
      Java for Everyone: Late Objects
    </title>
    ...
  </book>
```

If the author element had no content, then you *might* find this, or...

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20, 2011">
    <author />
    <title>
      Java for Everyone: Late Objects
    </title>
    ...
  </book>
```

... or you *might* find  
a single  
**self-closing tag**  
like this.

# Example XML Document/File

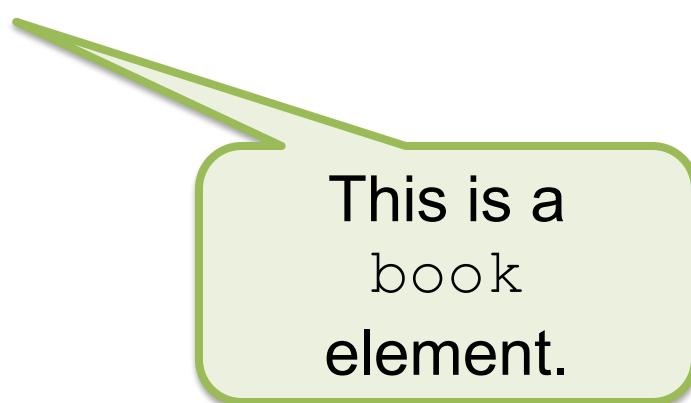
```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
  <author>Cay Horstmann</author>
  <title>
    Java for Everyone: Late Objects
  </title>
  ...
</book>
```



This is a  
title  
element.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
    <author>Cay Horstmann</author>
    <title>
        Java for Everyone: Late Objects
    </title>
    ...
</book>
```



This is a  
book  
element.

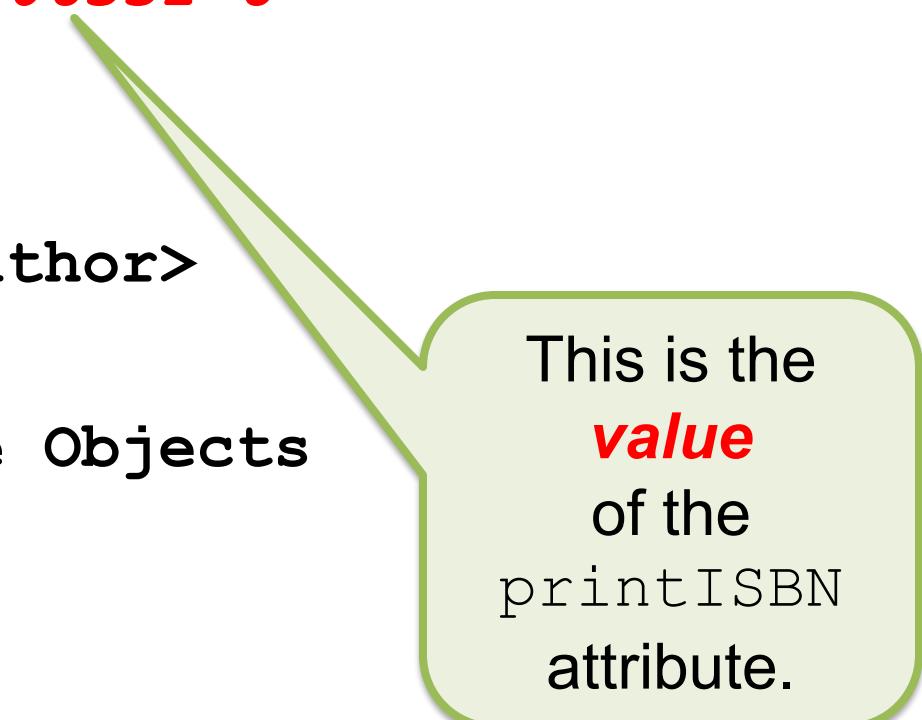
# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118-063-31-7"
      pubDate="Dec 20 2011"
      <author>Cay Horstmann</author>
      <title>
          Java for Everyone: Late Objects
      </title>
      ...
</book>
```

This is the  
**name** of an  
**attribute**  
of the book  
element.

# Example XML Document/File

```
<?xml version="1.0" encoding="UTF-8"?>
<book printISBN="978-1-118-06331-6"
      webISBN="1-118063-31-7"
      pubDate="Dec 20 2011">
  <author>Cay Horstmann</author>
  <title>
    Java for Everyone: Late Objects
  </title>
  ...
</book>
```



This is the  
**value**  
of the  
printISBN  
attribute.

# Recursive Structure

- An XML document (without the XML declaration in the first line) is made up of:
  - A ***top-level element***
  - A string of zero or more ***child elements*** of the top-level element, each of which is exactly like the top-level element of an XML document
- Notice the similarity to a tree: the structure of an XML document is also ***recursive***
- Information it represents is ***hierarchical***

# Can You Find All The Errors?

```
<?xml version="1.0" encoding="UTF-8"?>
<pony>
  <unicorn mark="three lozenges">
    <color>cyan</unicorn>
    </color>
    <unicorn>Twilight<color>cyan</color>Sparkle</unicorn>
    <unicorn mark="dolphins" version="G4" />
  </pony>
  <dragon>
    <youth>Spike
  </dragon>
```

# Resources

- *Big Java*, Section 23.1 (but not the rest of Chapter 23)
  - <https://learning.oreilly.com/library/view/big-java-4th/9780470509487/>
- Wikipedia: XML
  - <https://en.wikipedia.org/wiki/XML>