XML

Prof. P. M. Jadav,
Associate Professor,
CE Department,
DDU, Nadiad

Markup Language

A markup language must specify

- What markup is allowed
- What markup is required
- How markup is to be distinguished from text
- What the markup means

*XML only specify the first three, the fourth is specified by DTD

SGML(ISO 8879)

- Standard Generalized Markup Language
- The international standard for defining descriptions of structure and content in text documents
- Interchangeable: device-independent, system-independent
- tags are not predefined
- Using DTD to validate the structure of the document
- Large, powerful, and very complex
- Heavily used in industrial and commercial for over a decade

HTML(RFC 1866)

- HyperText Markup Language
- A small SGML application used on web (a DTD and a set of processing conventions)
- Can only use a predefined set of tags

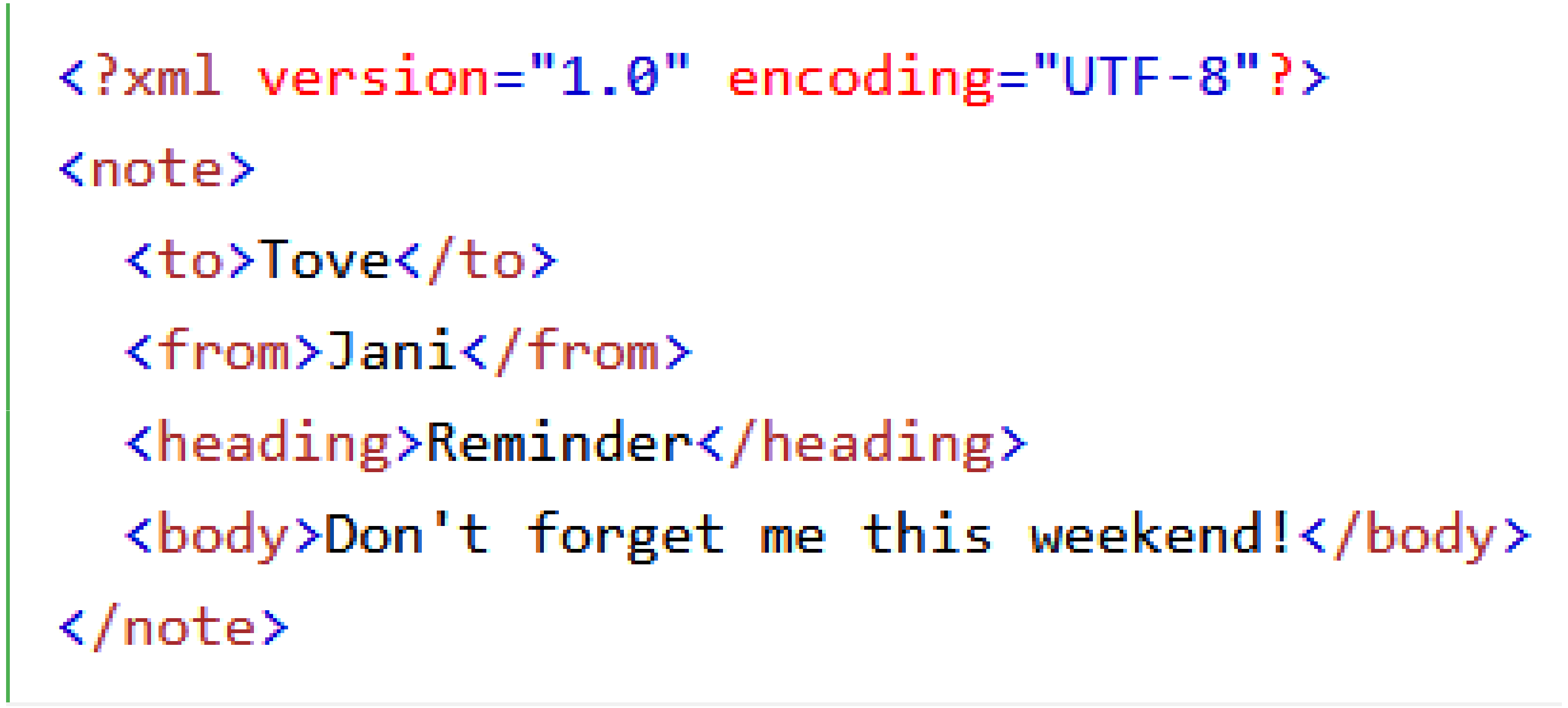
What is XML?

- stands for e**X**tensible **M**arkup **L**anguage.
- designed to store and transport data.
- designed to be both human- and machine-readable.
- is a markup language much like HTML
- was designed to be self-descriptive
- is a W3C Recommendation

What is XML?

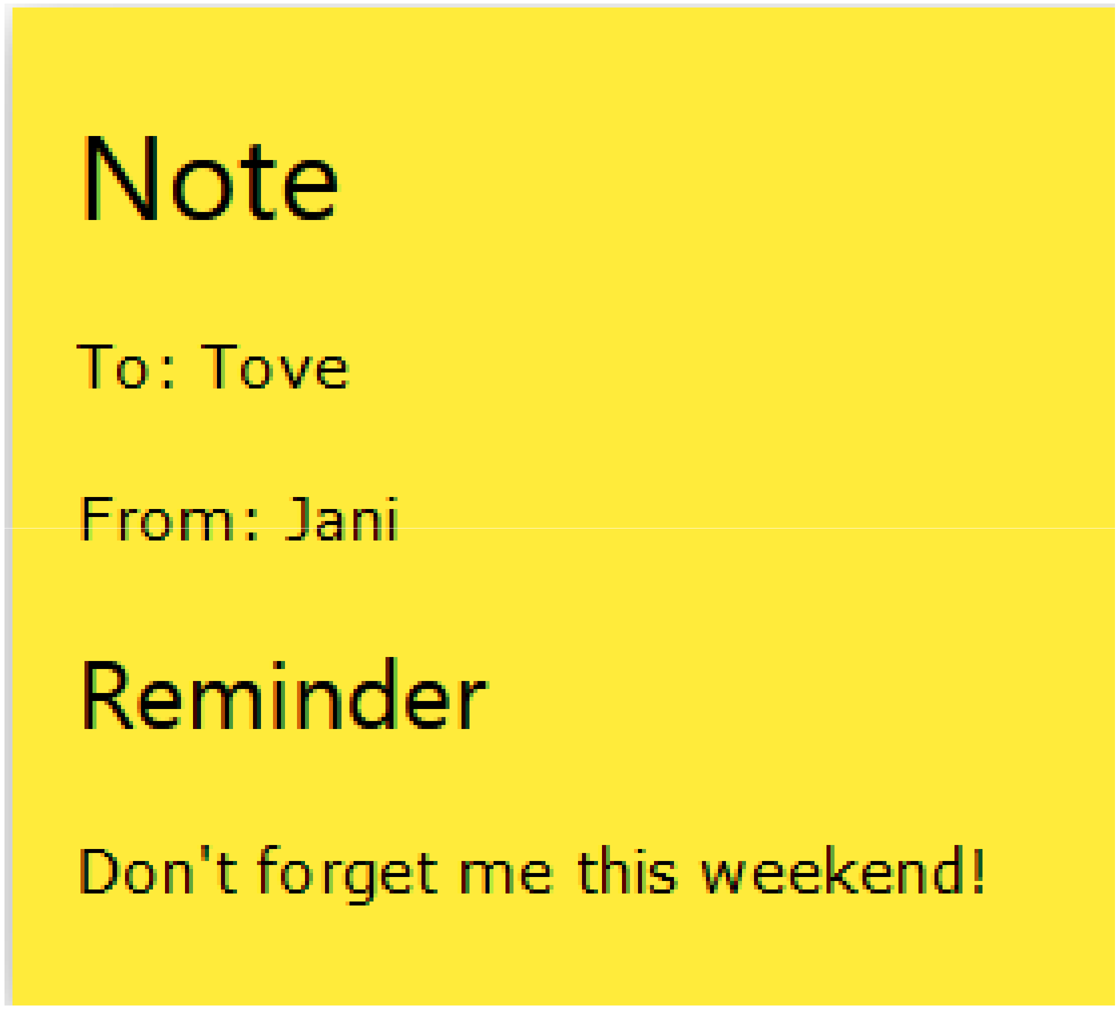
- A simplified version of SGML
- Maintains the most useful parts of SGML
- Designed so that SGML can be delivered over the Web
- More flexible and adaptable than HTML
- XHTML -- a reformulation of HTML 4 in XML 1.0

XML Example 1 (note.xml)



```xml
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

Someone must write a piece of software to send, receive, store, or display it:



Note

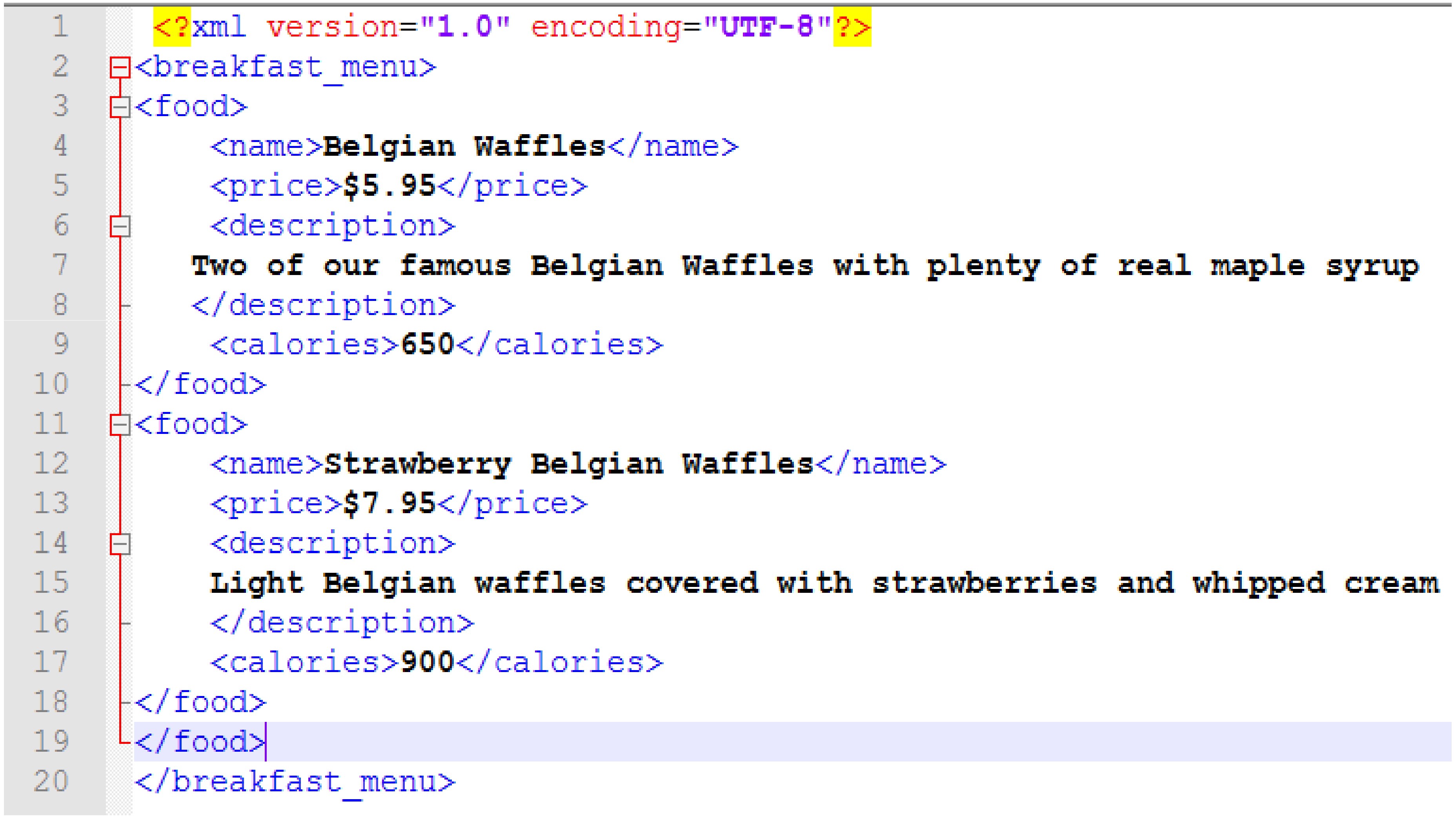
To: Tove

From: Jani

Reminder

Don't forget me this weekend!

XML Example 2 (foods.xml)



```
 <?xml version="1.0" encoding="UTF-8"?>
<breakfast_menu>
<food>
     <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>
     <name>Strawberry Belgian Waffles</name>
     <price>$7.95</price>
     <description>
     Light Belgian waffles covered with strawberries and whipped cream
     </description>
     <calories>900</calories>
</food>
</food>
</breakfast_menu>
```

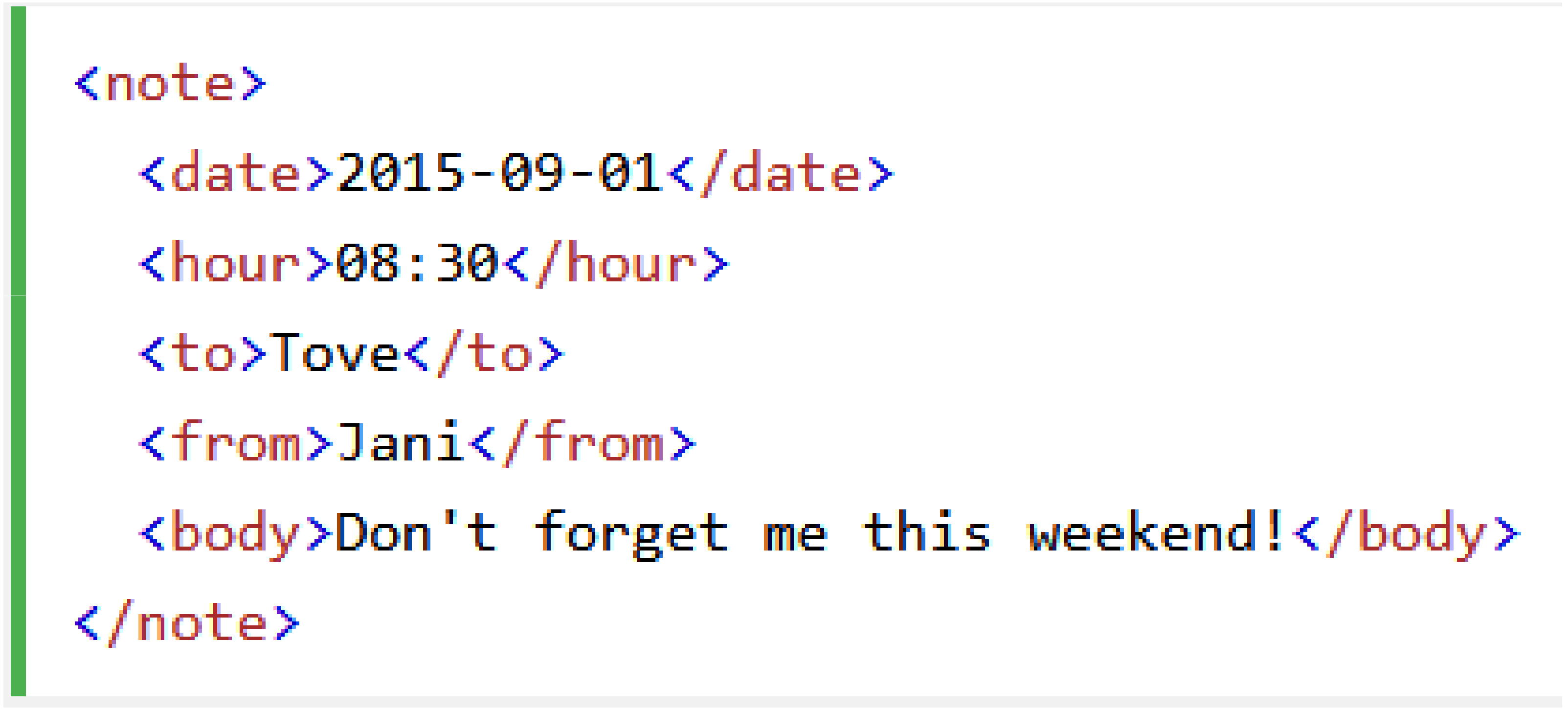
The Difference Between XML and HTML

- XML was designed to carry data - with focus on what data is
- HTML was designed to display data - with focus on how data looks
- XML tags are not predefined like HTML tags are
- XML is not a replacement for HTML

XML Does Not Use Predefined Tags

- <to> and <from> are not defined in any XML standard. These tags are "invented" by the author of the XML document.
- HTML works with predefined tags like <p>, <h1>, <table>, etc.
- With XML, the author must define both the tags and the document structure.

XML is Extensible (note_new.xml)



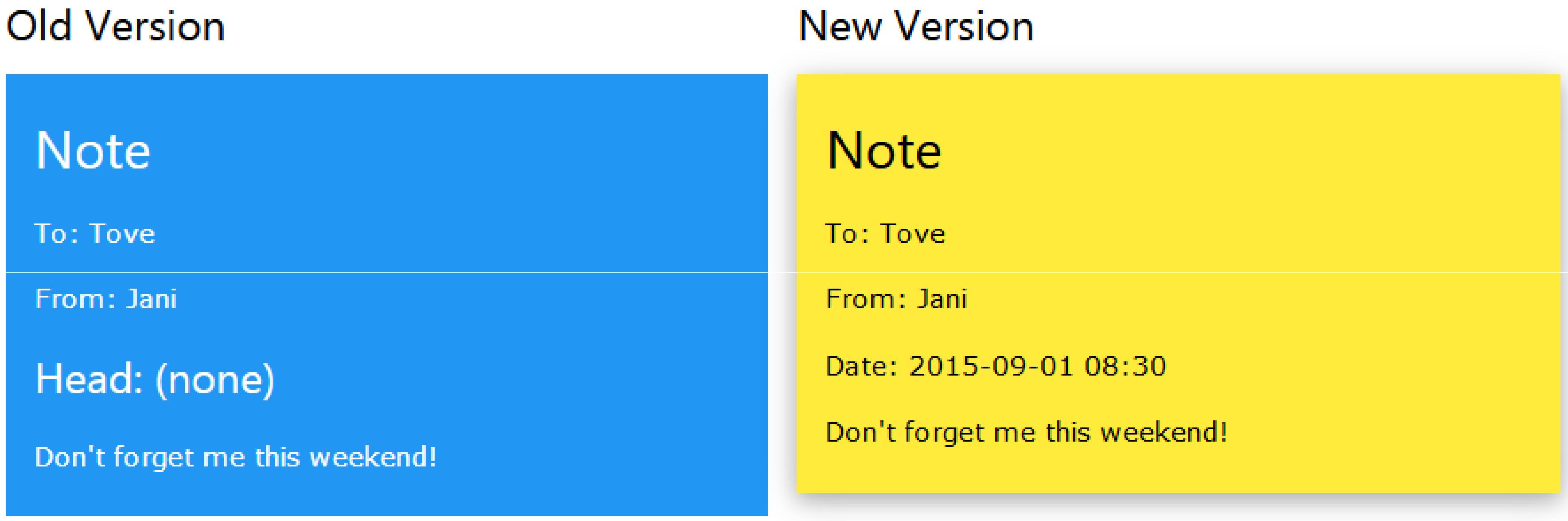
```xml
<note>
  <date>2015-09-01</date>
  <hour>08:30</hour>
  <to>Tove</to>
  <from>Jani</from>
  <body>Don't forget me this weekend!</body>
</note>
```

note.xml

```xml
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

note_new.xml

```xml
<note>
  <date>2015-09-01</date>
  <hour>08:30</hour>
  <to>Tove</to>
  <from>Jani</from>
  <body>Don't forget me this weekend!</body>
</note>
```



Old Version

Note

To: Tove

From: Jani

Head: (none)

Don't forget me this weekend!

New Version

Note

To: Tove

From: Jani

Date: 2015-09-01 08:30

Don't forget me this weekend!

XML Simplifies Things

It simplifies

- data sharing
- data transport
- platform changes
- data availability

How to use XML?

- **XML Separates Data from Presentation**
  - XML does not carry any information about how to be displayed
  - The same XML data can be used in many different presentation scenarios.
- **XML is Often a Complement to HTML**
  - XML is used to store or transport data, while HTML is used to format and display the same data.

Why is XML Important?

- Plain Text
  - Easy to edit
  - Useful for storing small amounts of data
  - Possible to efficiently store large amounts of XML data through an XML front end to a database

- Data Identification
  - Tell you what kind of data you have
  - Can be used in different ways by different applications

Why is XML Important?

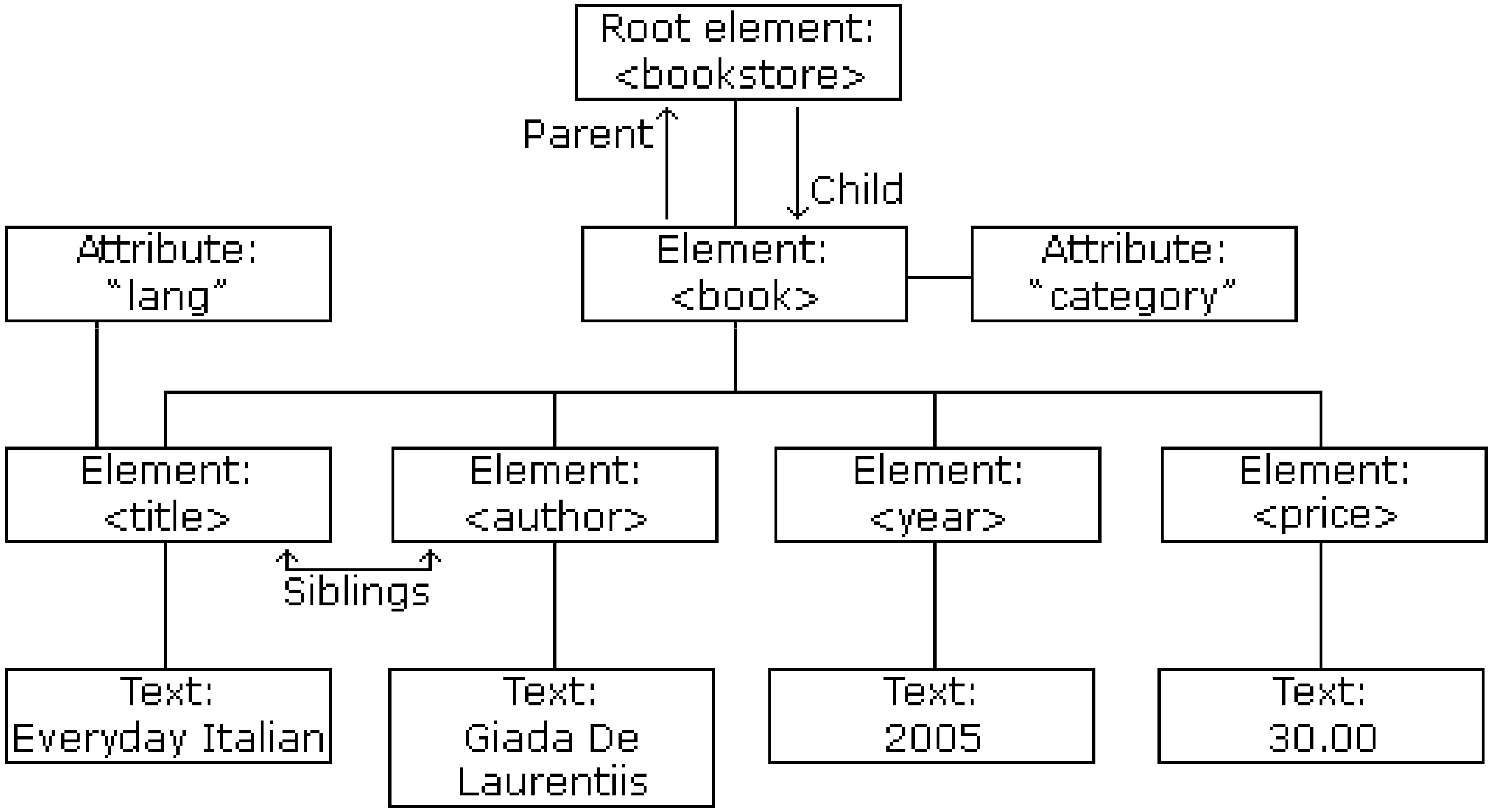
- Stylability
  - Inherently style-free
  - XSL---Extensible Stylesheet Language
  - Different XSL formats can then be used to display the same data in different ways

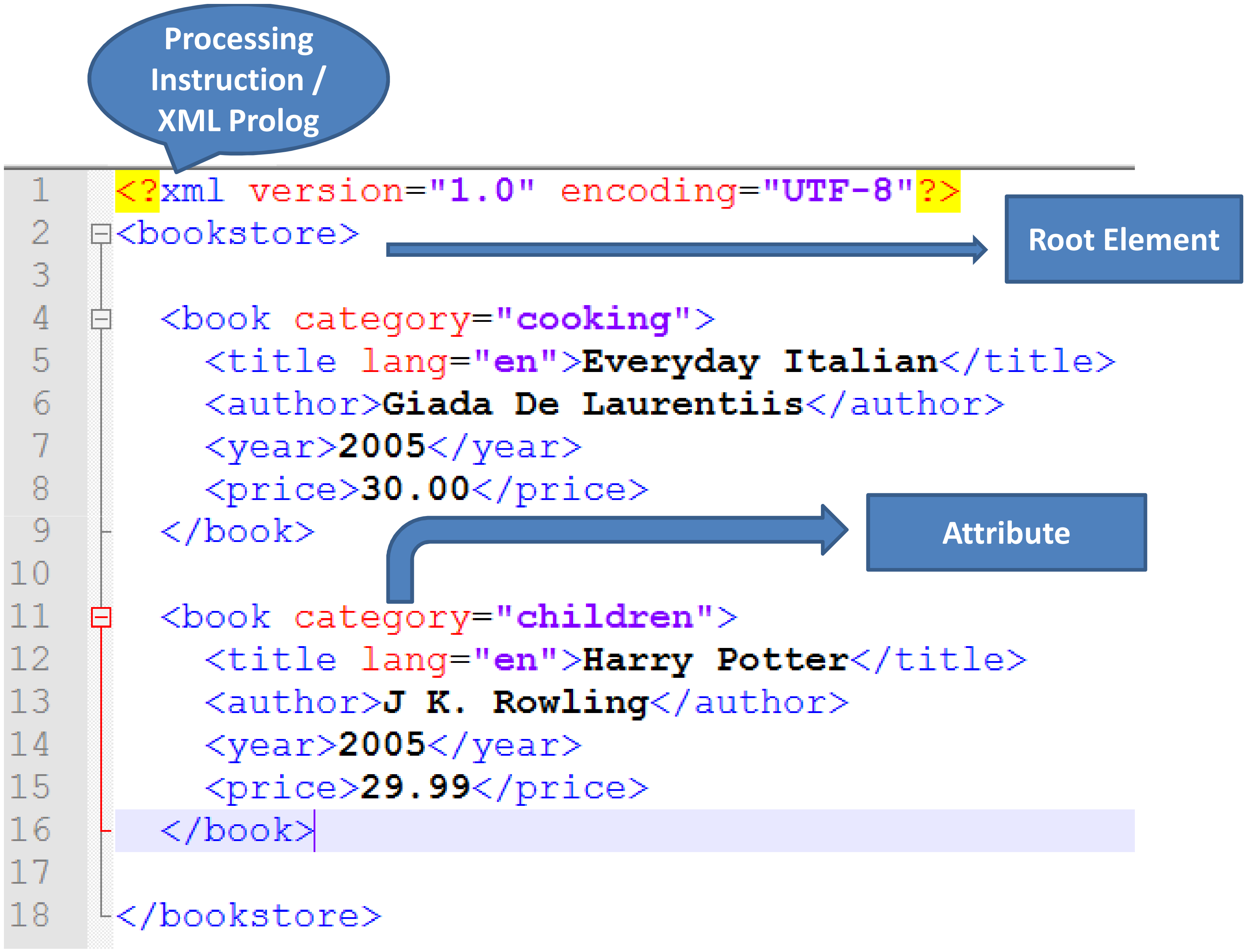
- Inline Reusabiliy
  - Can be composed from separate entities
  - Modularize your documents without resorting to links

Why is XML Important?

- Linkability -- XLink and XPointer
  - Simple unidirectional hyperlinks
  - Two-way links
  - Multiple-target links
  - “Expanding” links
- Easily Processed
  - Regular and consistent notation
  - Vendor-neutral standard
- Hierarchical
  - Faster to access
  - Easier to rearrange

XML Tree Structure



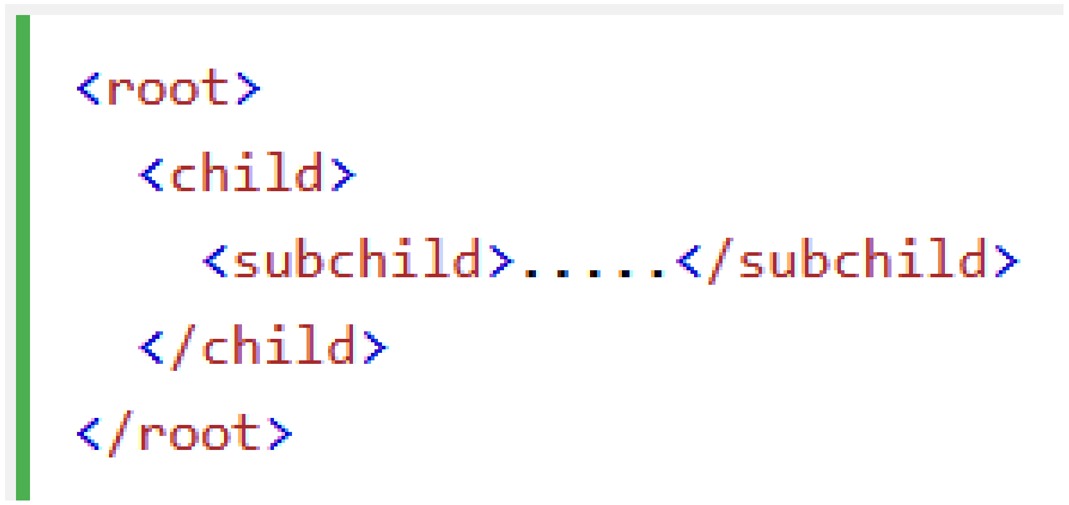


```xml
<?xml version="1.0" encoding="UTF-8"?>
<bookstore>

  <book category="cooking">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>

  <book category="children">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>

</bookstore>
```

XML Syntax Rules (Well Formed XML)

1) XML Documents Must Have a Root Element

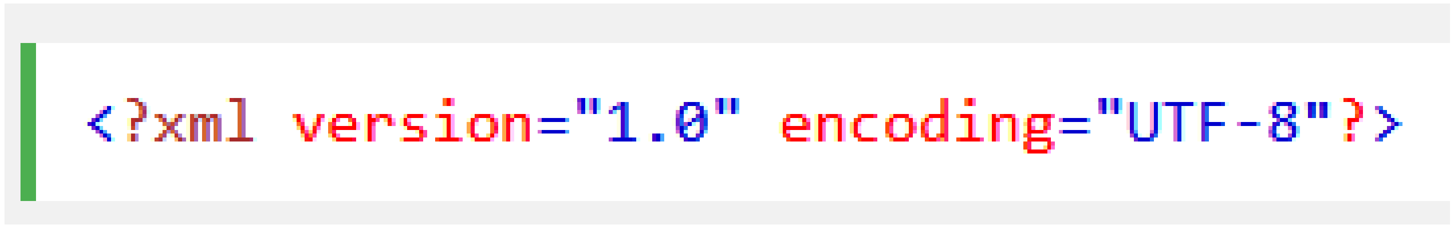
– XML documents must contain one **root** element that is the **parent** of all other elements:



```
<root>
  <child>
    <subchild>.....</subchild>
  </child>
</root>
```

XML Syntax Rules

2) The XML Prolog

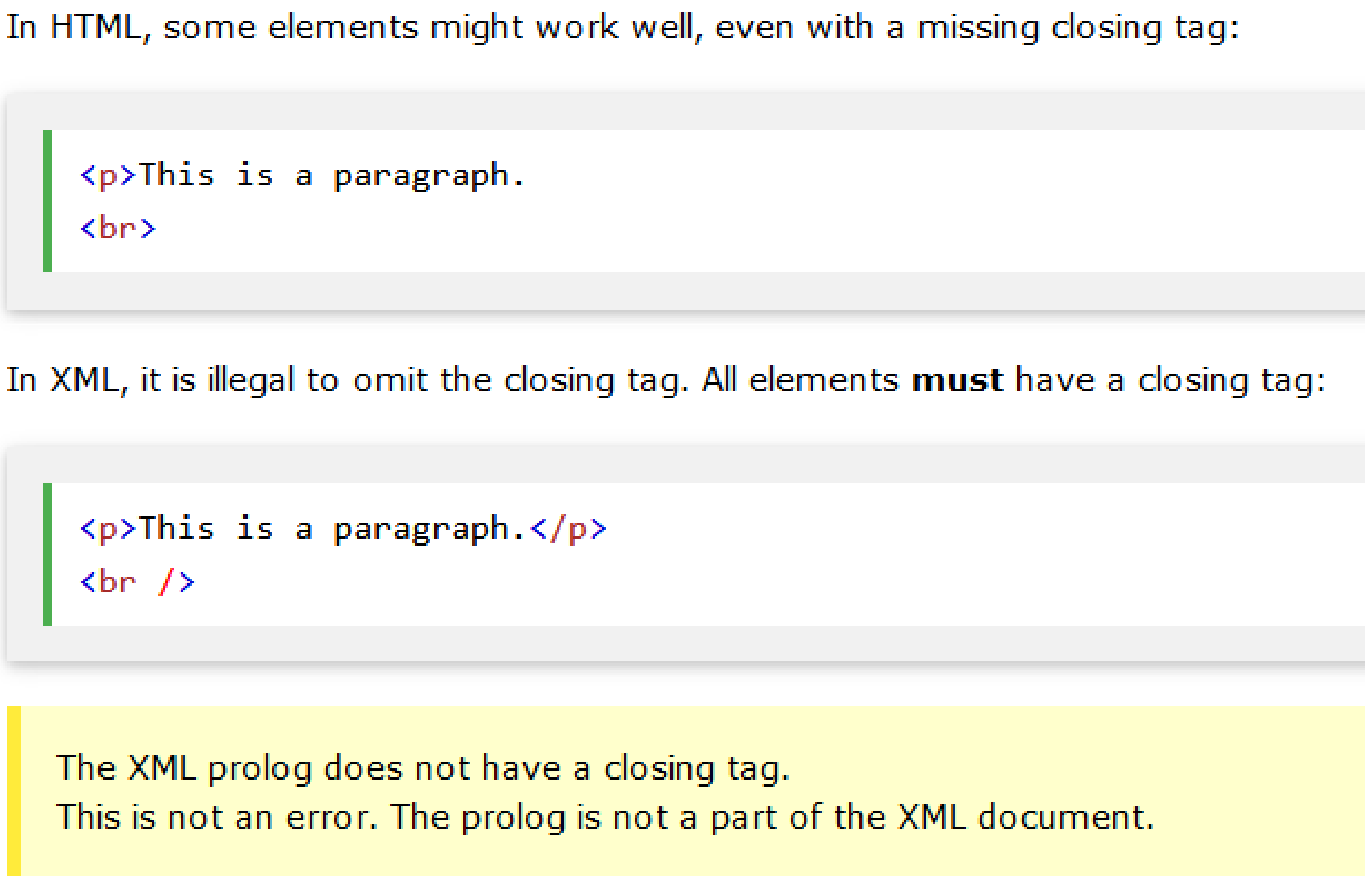


```xml
<?xml version="1.0" encoding="UTF-8"?>
```

- The XML prolog is optional. If it exists, it must come first in the document
- XML documents can contain international chars
- To avoid errors, you should specify the encoding used, or save your XML files as UTF-8
- UTF-8 is the default char. encoding for XML docs

XML Syntax Rules

3) All XML Elements Must Have a Closing Tag



In HTML, some elements might work well, even with a missing closing tag:

```
<p>This is a paragraph.
<br>
```

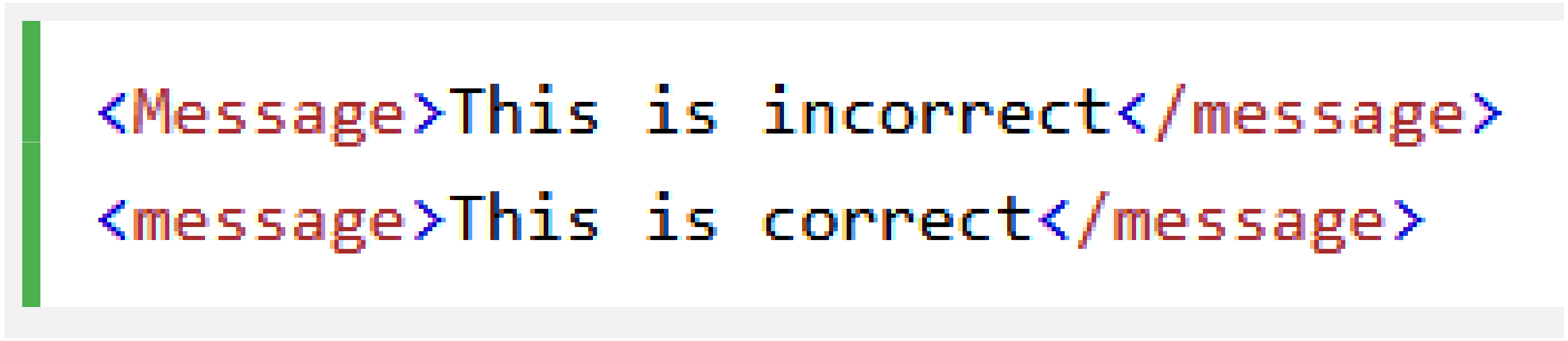
In XML, it is illegal to omit the closing tag. All elements **must** have a closing tag:

```
<p>This is a paragraph.</p>
<br />
```

The XML prolog does not have a closing tag.
This is not an error. The prolog is not a part of the XML document.

XML Syntax Rules

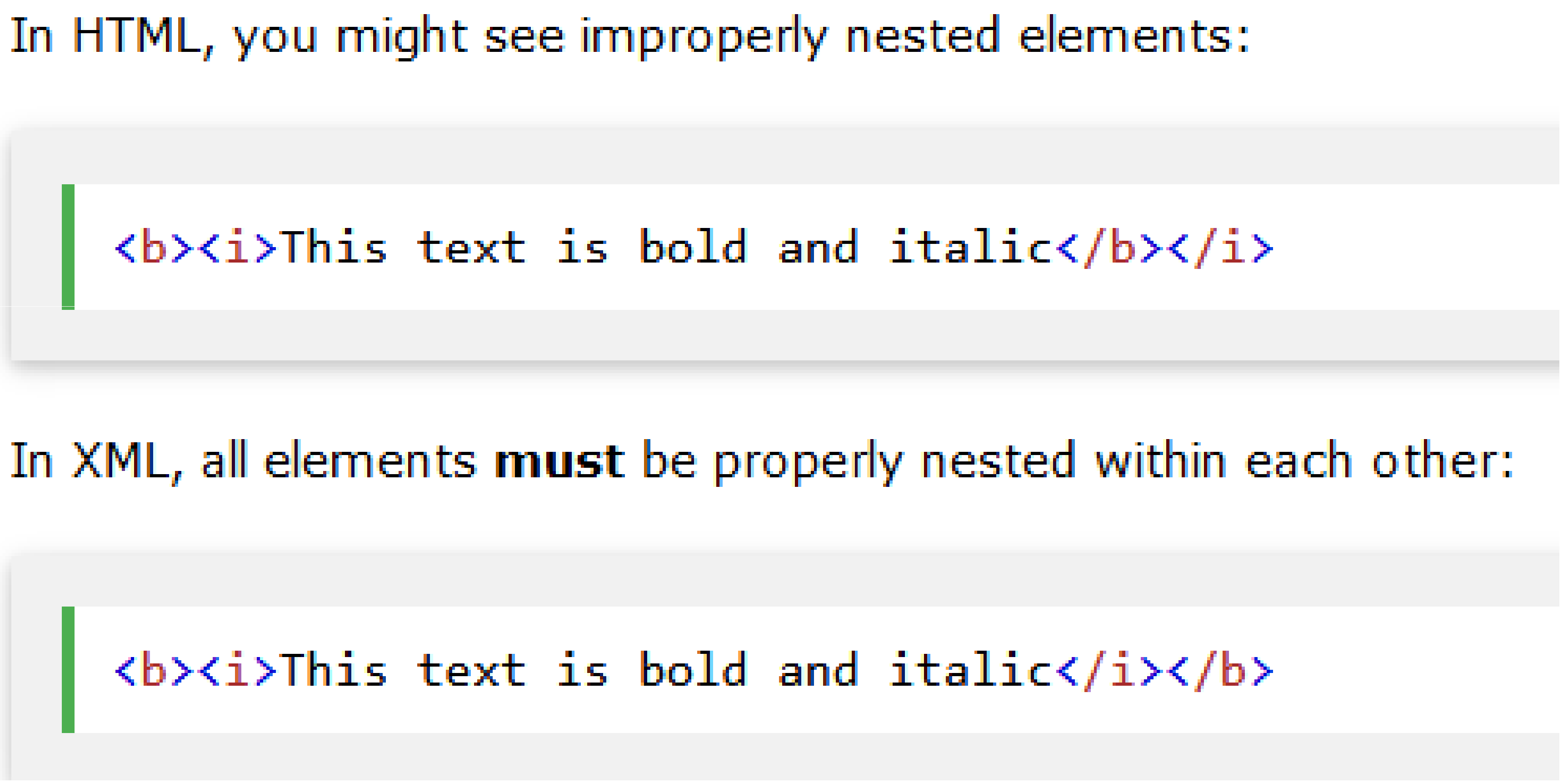
4) XML Tags are Case Sensitive



```
<Message>This is incorrect</message>
<message>This is correct</message>
```

XML Syntax Rules

5) XML Elements Must be Properly Nested



In HTML, you might see improperly nested elements:

```
<b><i>This text is bold and italic</b></i>
```

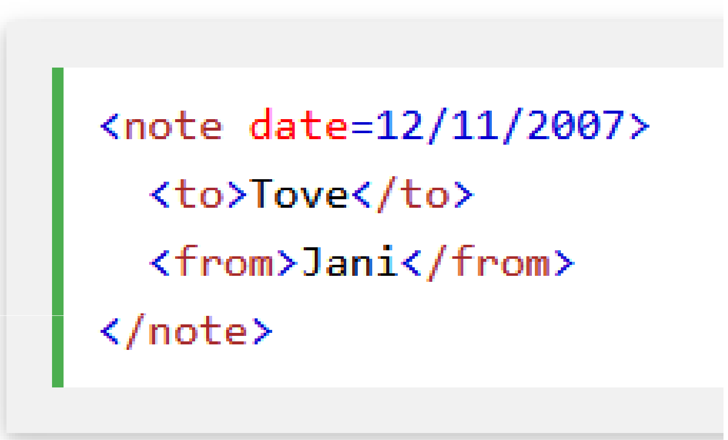
In XML, all elements **must** be properly nested within each other:

```
<b><i>This text is bold and italic</i></b>
```

XML Syntax Rules

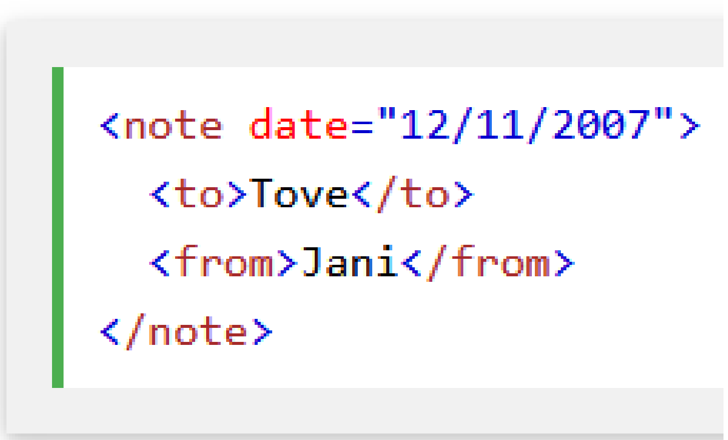
6) XML Attribute Values Must be Quoted

INCORRECT:



```
<note date=12/11/2007>
  <to>Tove</to>
  <from>Jani</from>
</note>
```

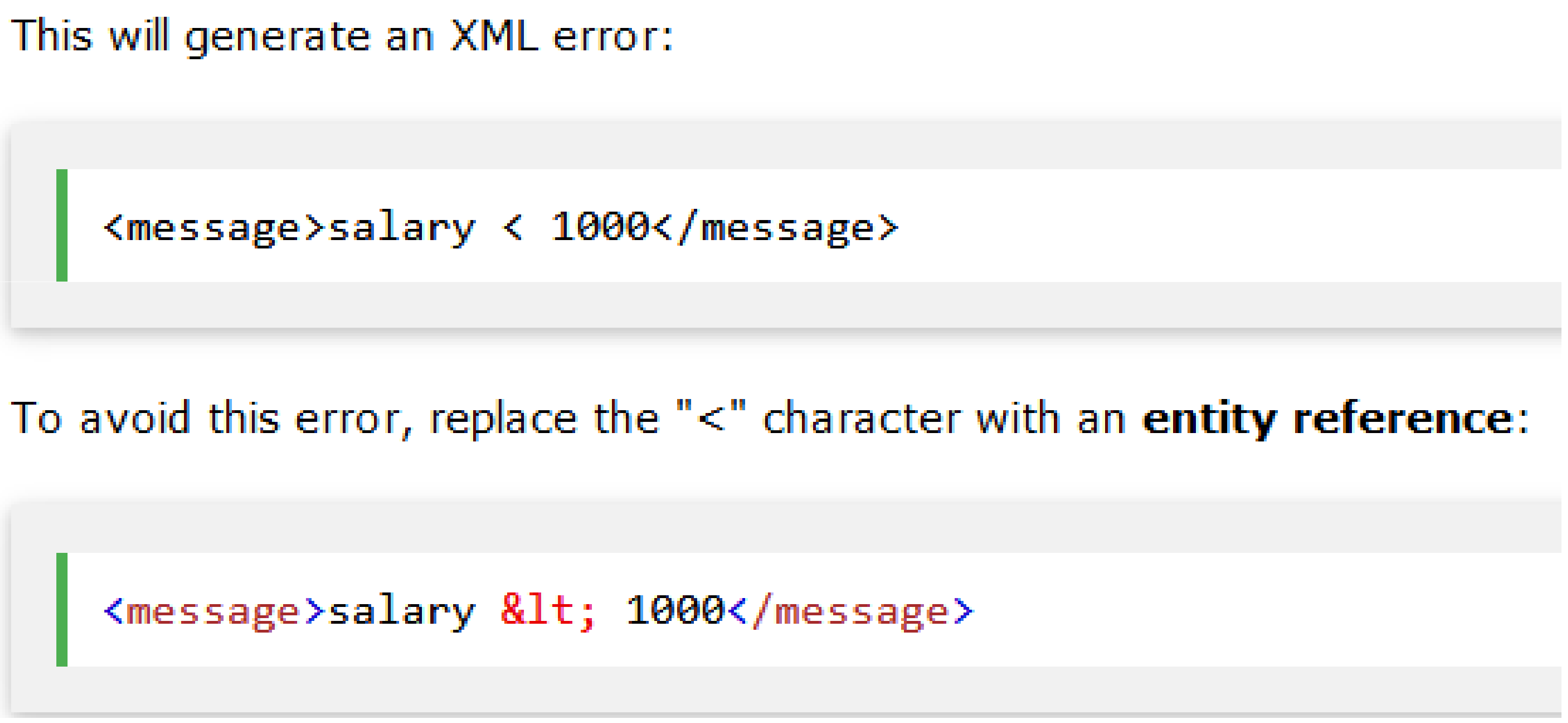
CORRECT:



```
<note date="12/11/2007">
  <to>Tove</to>
  <from>Jani</from>
</note>
```

XML Syntax Rules

7) Entity References



This will generate an XML error:

```
<message>salary < 1000</message>
```

To avoid this error, replace the "<" character with an **entity reference**:

```
<message>salary &lt; 1000</message>
```

XML Syntax Rules

7) Entity References

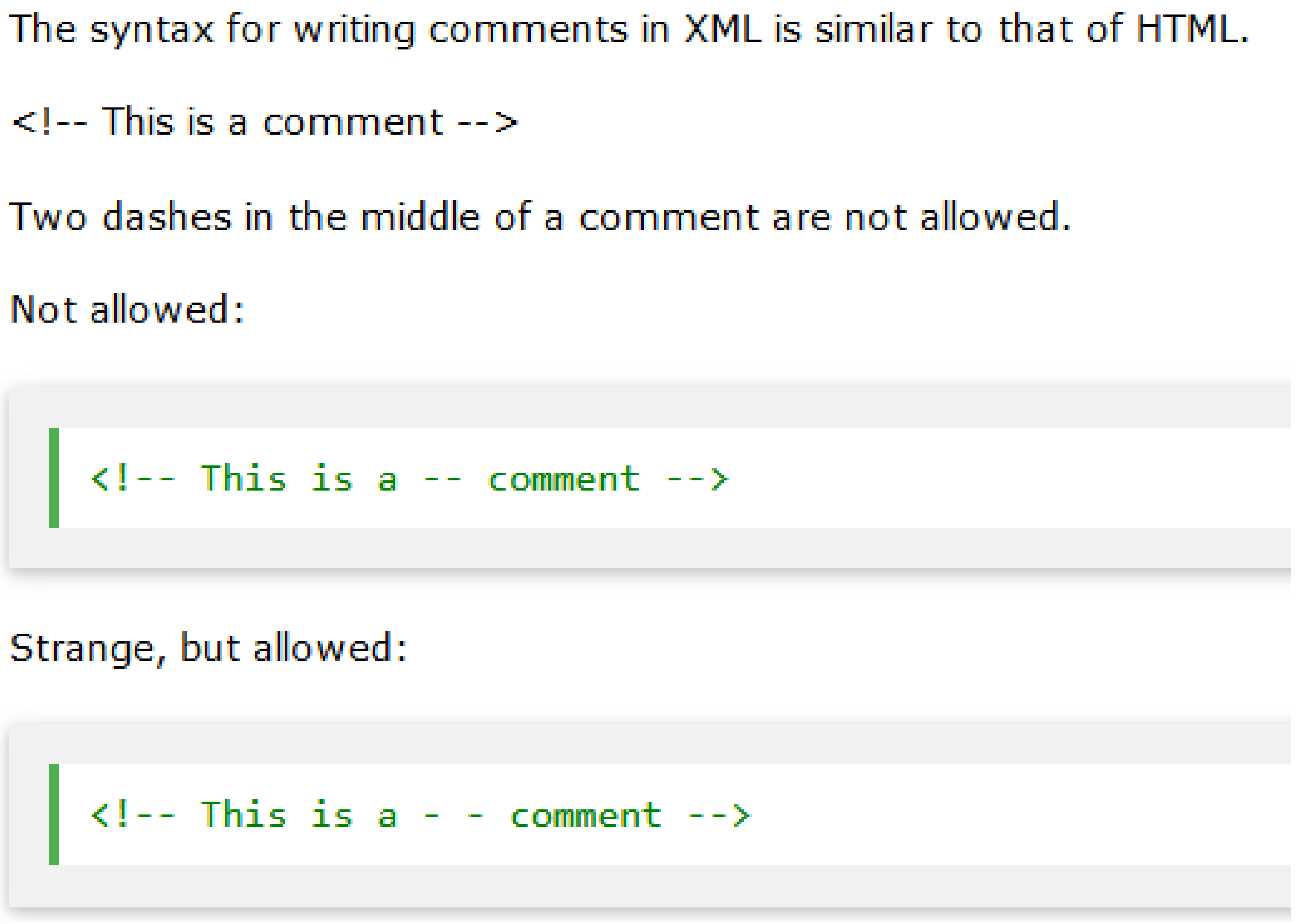
There are 5 pre-defined entity references in XML:

&lt;	&lt;	less than
&gt;	&gt;	greater than
&amp;	&amp;	ampersand
&apos;	'	apostrophe
&quot;	"	quotation mark

Only &lt; and &amp; are strictly illegal in XML, but it is a good habit to replace &gt; with &gt; as well.

XML Syntax Rules

8) Comments in XML



The syntax for writing comments in XML is similar to that of HTML.

<!-- This is a comment -->

Two dashes in the middle of a comment are not allowed.

Not allowed:

```
<!-- This is a -- comment -->
```

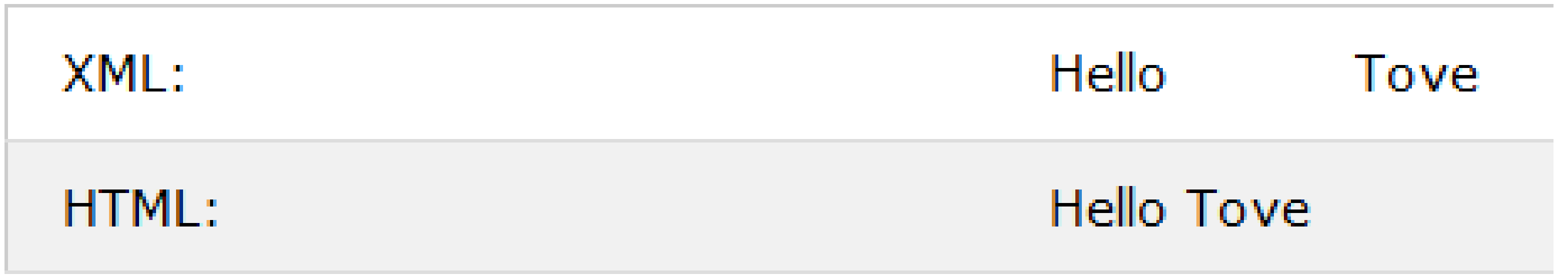
Strange, but allowed:

```
<!-- This is a - - comment -->
```

XML Syntax Rules

9) White-space is Preserved in XML

- XML does not truncate multiple white-spaces
- HTML truncates multiple white-spaces to one single white-space



XML:	Hello Tove
HTML:	Hello Tove

XML Syntax Rules

10) XML Stores New Line as LF

– Windows applications store a new line as:
  carriage return and line feed (CR+LF)

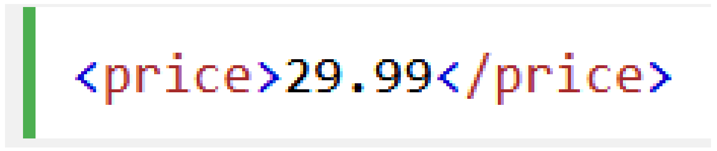
– Unix and Mac OSX uses LF

– Old Mac systems uses CR

– XML stores a new line as LF.

XML Elements

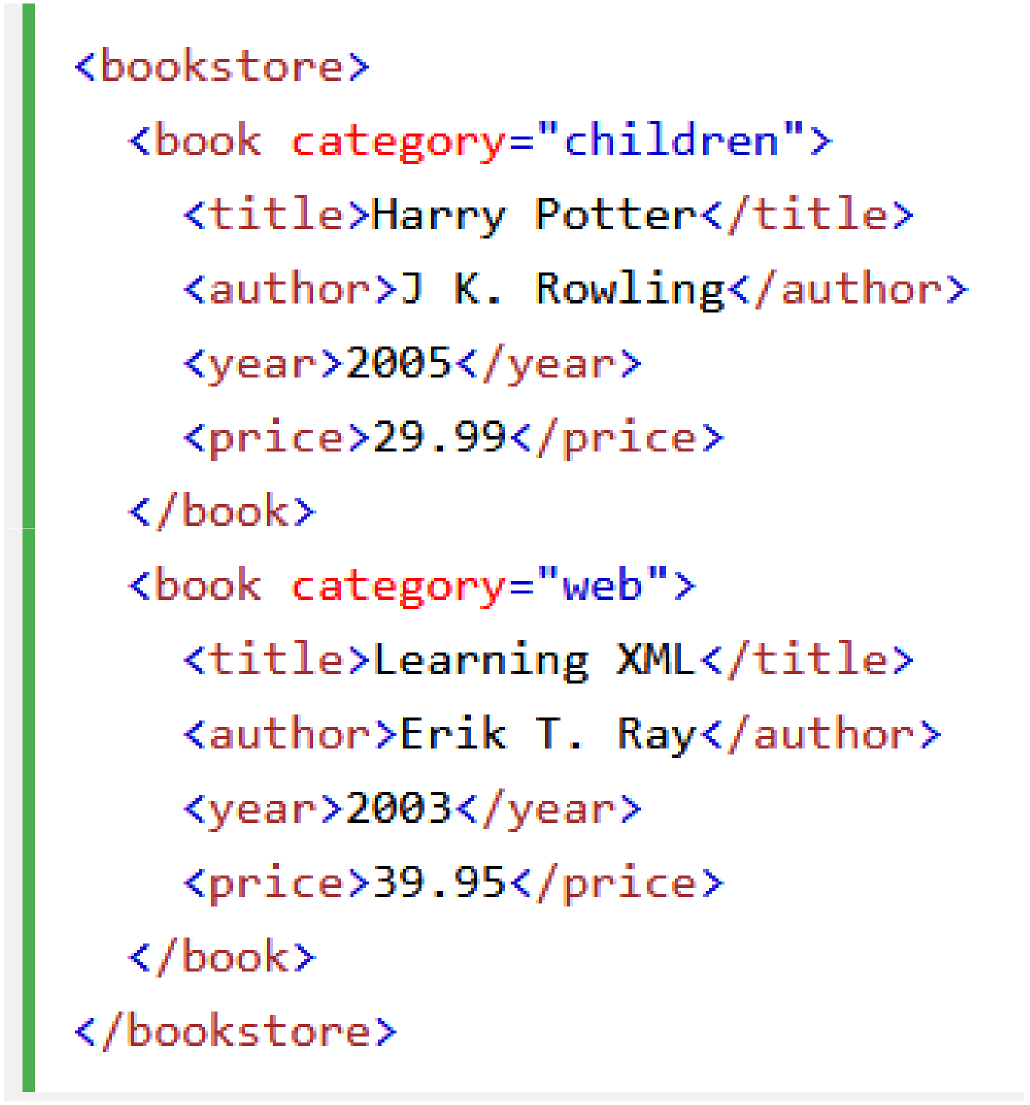
- An XML element is everything from (including) the element's start tag to (including) the element's end tag.



```
<price>29.99</price>
```

- An element can contain:
  - text
  - attributes
  - other elements
  - or a mix of the above

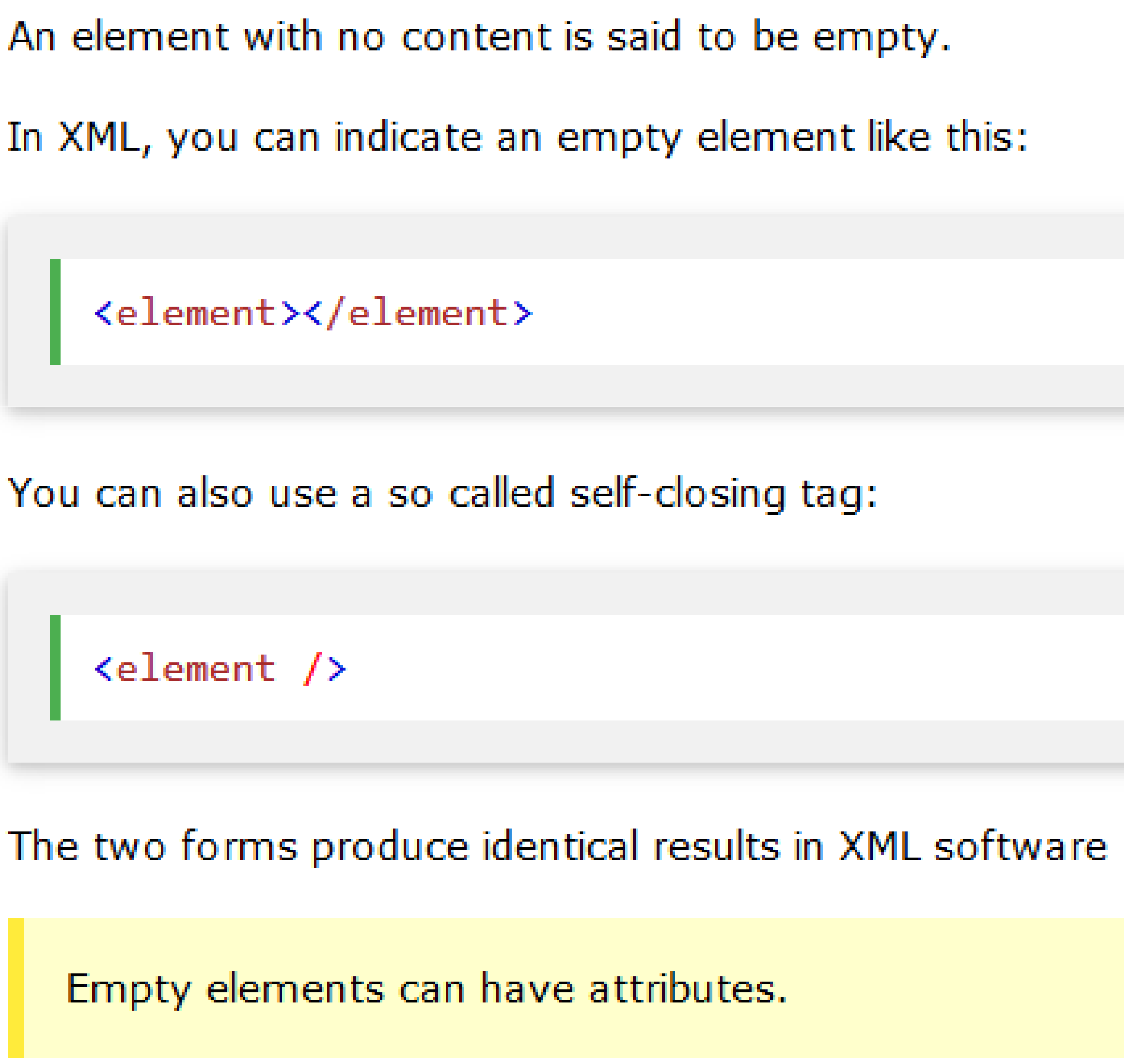
XML Elements



```xml
<bookstore>
  <book category="children">
    <title>Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="web">
    <title>Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
```

XML Elements

- **Empty XML Elements**



An element with no content is said to be empty.

In XML, you can indicate an empty element like this:

```
<element></element>
```

You can also use a so called self-closing tag:

```
<element />
```

The two forms produce identical results in XML software

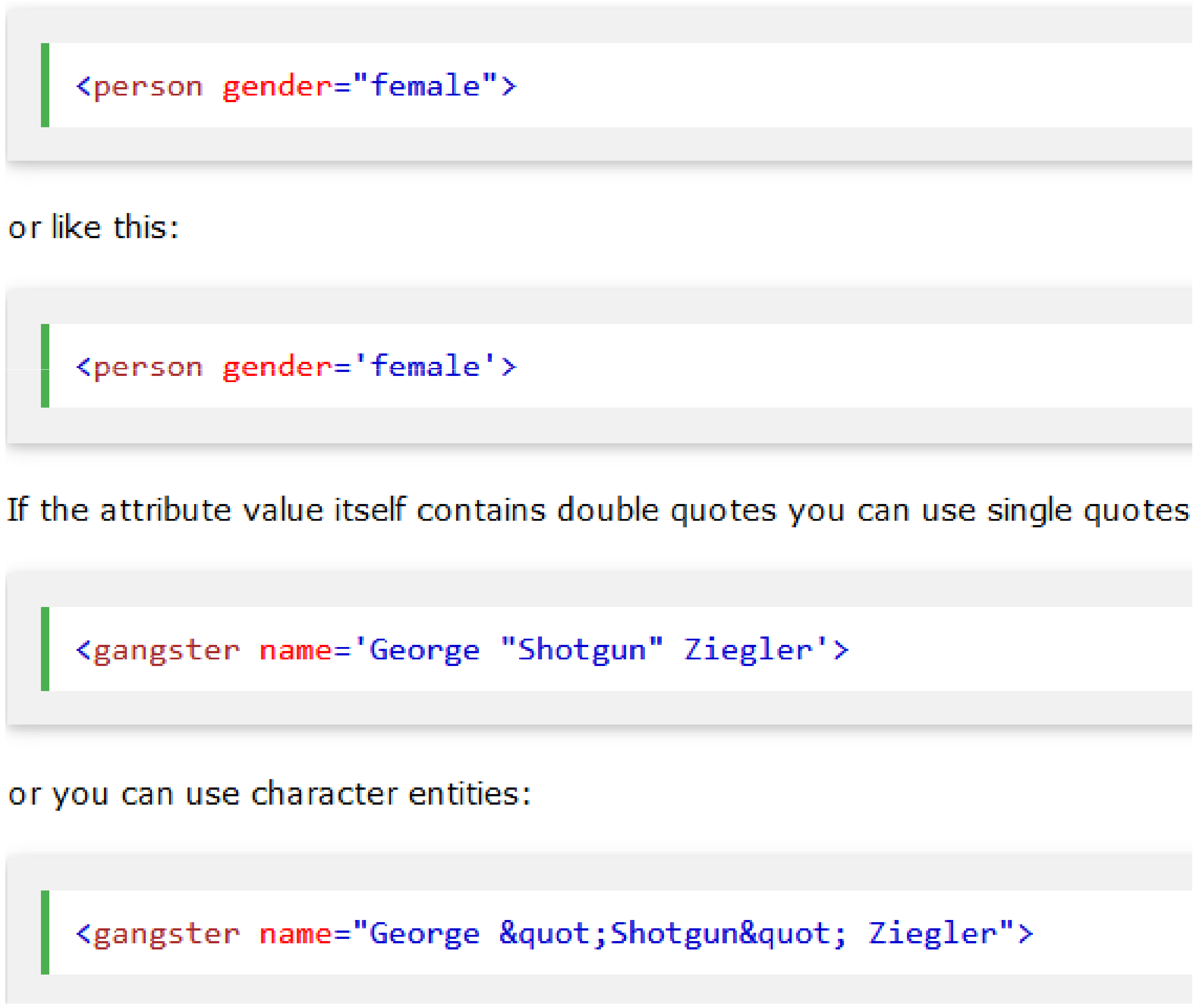
Empty elements can have attributes.

XML Elements

- **XML Naming Rules**
  - Element names are case-sensitive
  - Element names must start with a letter or underscore
  - Element names cannot start with the letters *xml* (or XML, or Xml, etc)
  - Element names can contain letters, digits, hyphens, underscores, and periods
  - Element names cannot contain spaces
  - Any name can be used, no words are reserved (except xml).

XML Attributes

- **XML Attributes Must be Quoted**



```
<person gender="female">
```

or like this:

```
<person gender='female'>
```

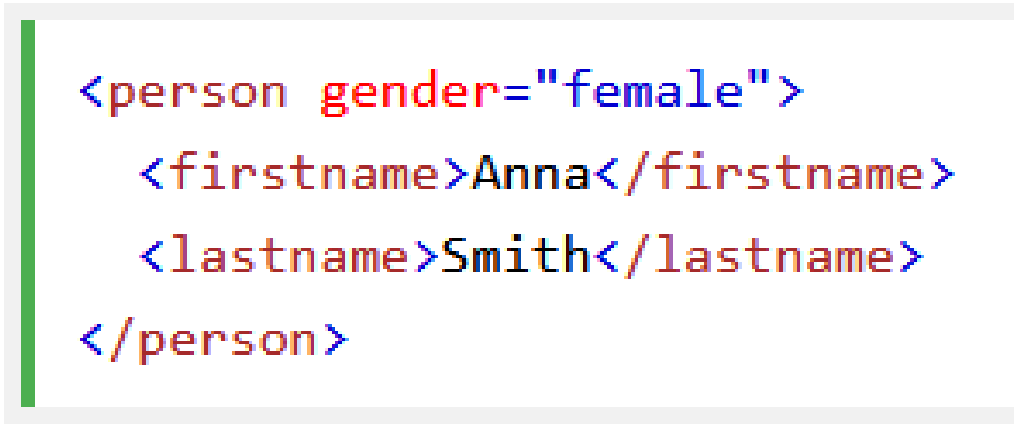
If the attribute value itself contains double quotes you can use single quotes

```
<gangster name='George "Shotgun" Ziegler'>
```

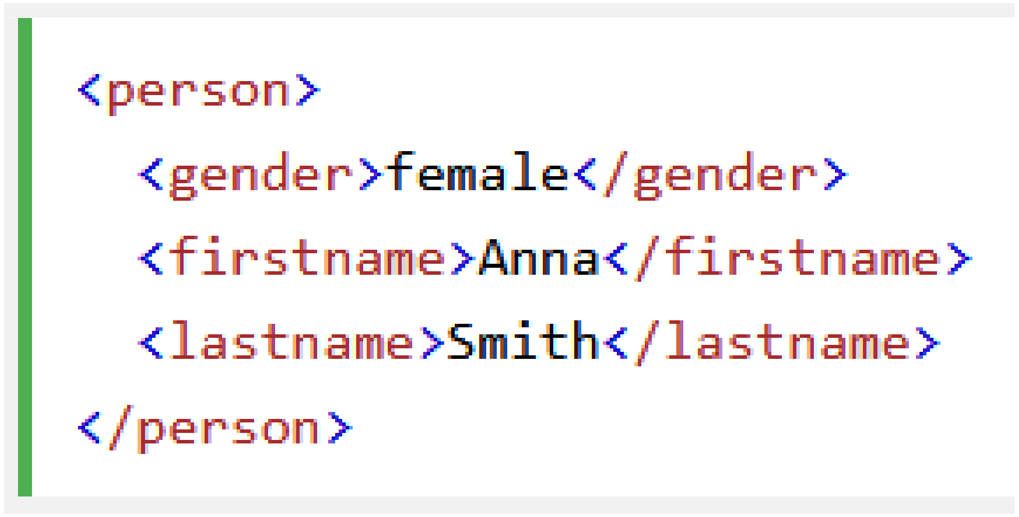
or you can use character entities:

```
<gangster name="George &quot;Shotgun&quot; Ziegler">
```

XML Elements vs. Attributes

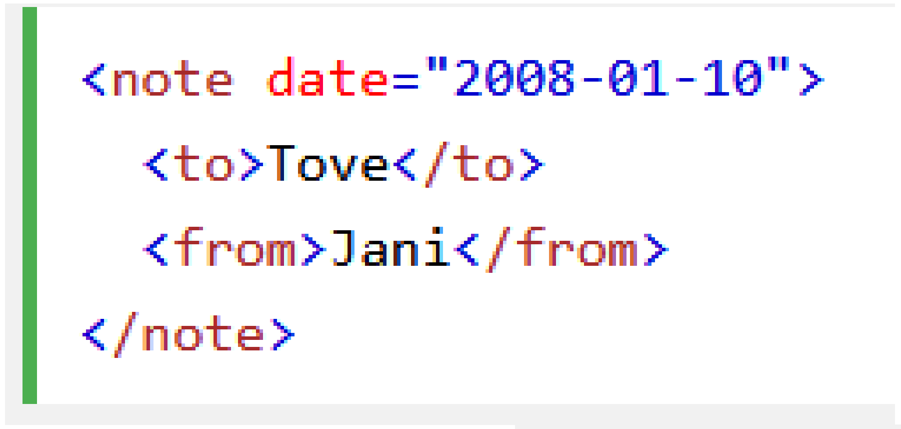


```xml
<person gender="female">
  <firstname>Anna</firstname>
  <lastname>Smith</lastname>
</person>
```

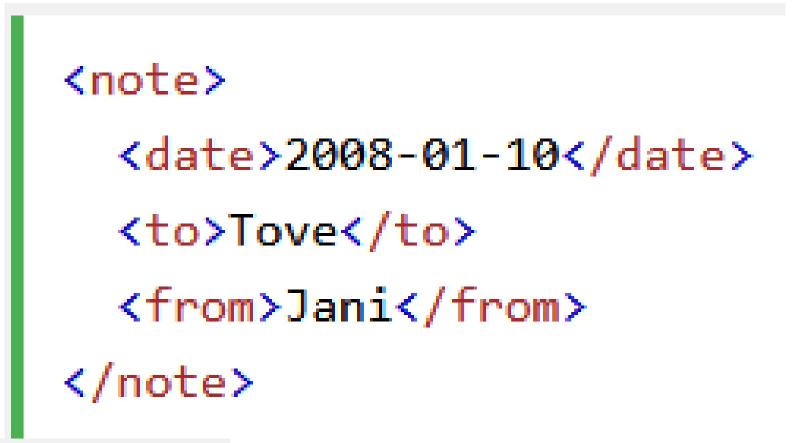


```xml
<person>
  <gender>female</gender>
  <firstname>Anna</firstname>
  <lastname>Smith</lastname>
</person>
```

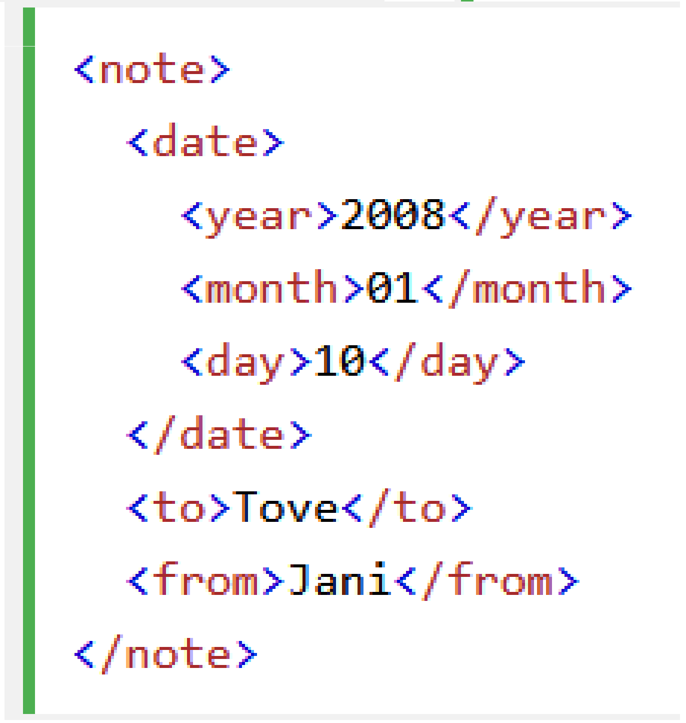
XML Elements vs. Attributes



```xml
<note date="2008-01-10">
  <to>Tove</to>
  <from>Jani</from>
</note>
```



```xml
<note>
  <date>2008-01-10</date>
  <to>Tove</to>
  <from>Jani</from>
</note>
```



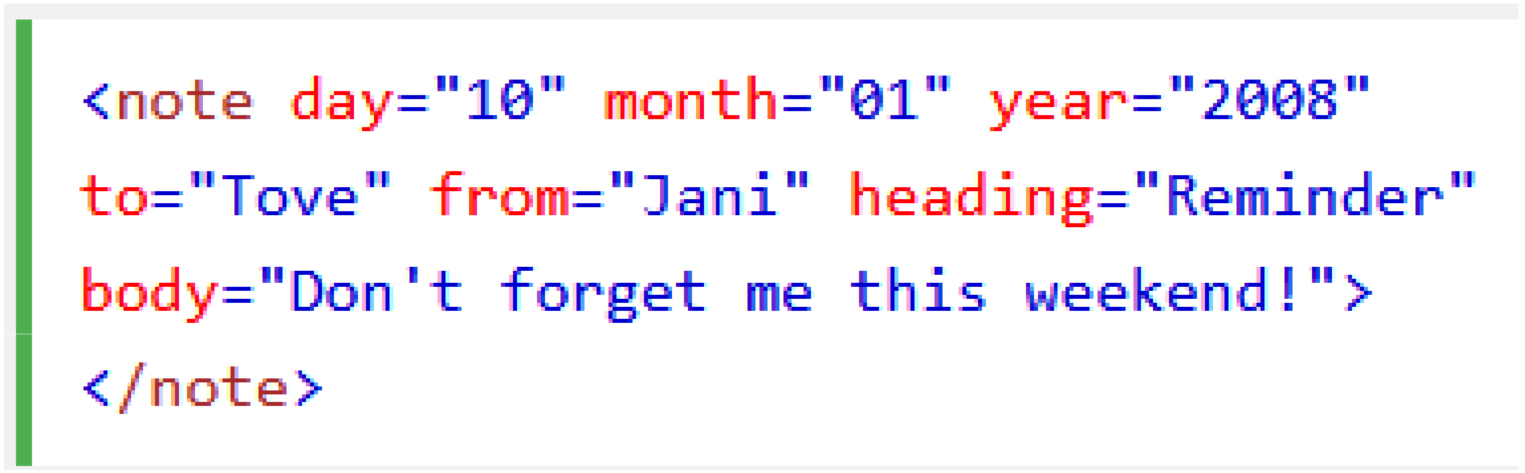
```xml
<note>
  <date>
    <year>2008</year>
    <month>01</month>
    <day>10</day>
  </date>
  <to>Tove</to>
  <from>Jani</from>
</note>
```

Should we avoid XML Attributes?

- Attributes
  - cannot contain multiple values (elements can)
  - cannot contain tree structures (elements can)
  - are not easily expandable (for future changes)

Should we avoid XML Attributes?

- Don't end up like this:



```
<note day="10" month="01" year="2008"
to="Tove" from="Jani" heading="Reminder"
body="Don't forget me this weekend!">
</note>
```

- HINT: metadata (data about data) should be stored as attributes, and that data itself should be stored as elements

XML Application1

XML can Separate Data from HTML

- Store data in separate XML files
- Using HTML for layout and display
- Using Data Islands
- Data Islands can be bound to HTML elements

Benefits:

Changes in the underlying data will not require any changes to your HTML

XML Application2

XML is used to Exchange Data

- Text format
- Software-independent, hardware-independent
- Exchange data between incompatible systems, given that they agree on the same tag definition.
- Can be read by many different types of applications

Benefits:

- Reduce the complexity of interpreting data
- Easier to expand and upgrade a system

XML Application3

XML can be used to Store Data

- Plain text file
- Store data in files or databases
- Application can be written to store and retrieve information from the store
- Other clients and applications can access your XML files as data sources

Benefits:

Accessible to more applications

XML Application4

XML can be used to Create new Languages

- WML (Wireless Markup Language) used to markup Internet applications for handheld devices like mobile phones (WAP)
- MusicXML used to publishing musical scores

Java APIs for XML

- JAXP: Java API for XML Processing
- JAXB: Java Architecture for XML Binding
- JDOM: Java DOM
- DOM4J: an alternative to JDOM
- JAXM: Java API for XML Messaging (asynchronous)
- JAX-RPC: Java API for XML-based Remote Process Communications (synchronous)
- JAXR: Java API for XML Registries

References

- https://www.w3schools.com/xml/