9.2: The XML Document

In this section, we will look at an example XML document and describe its individual parts.

The XML document (dvd.xml) below is for an online DVD rental business. It contains only two DVDs titles for now, but it will be at the heart of an ambitious company's web application with hopes for very rapid growth.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE DVDS SYSTEM "dvd.dtd">
<DVDS>
  <DVD id="1">
   <title>Witness</title>
   <performers>
    <performer>Harrison  Ford</performer>
    <performer>Kelly  McGillis</performer>
   </performers>
   <discs>1</discs>
   <price>7.99</price>
  </DVD>
  <DVD id="2">
   <title>Star Wars Episode V - The  Empire Strikes Back</title>
   <performers>
    <performer>Mark  Hamill</performer>
    <performer>Harrison  Ford</performer>
    <performer>Carrie  Fisher</performer>
   </performers>
   <discs>2</discs>
   <price>14.95</price>
  </DVD>
</DVDS>
```

The first line of an XML document is the prolog. It contains the version, encoding information and the DTD (we will discuss DTDs in the next section), in this case dvd.dtd.

The prolog also specifies the root element, DVDS. All other elements are children of the root element.

After the prolog come the XML instance data and the beginning of the tree. It is essentially a collection of elements, starting with the root element. An element is comprised of the following:

- start tag
- attributes
- the data (either text or more elements)
- end tag

Tags consist of a word surrounded by angle brackets. For example, <DVD>.

An end tag is the same as the start tag but has a backslash in front of the name. Think of the start tag and end tag as parentheses to data.

Because XML documents are tree based, elements can be nested. In other words, elements can contain other elements (see lines 6 and 15).

Also, the start tag can have attributes. Attributes further describe the data and follow the format of: name="value" (see lines 4 and 13).

With all of these parts, it is critical to start and end elements correctly. An XML document that is properly written and formatted is called a **well-formed document**. A well-formed XML document makes it easier for an application to parse it.