Return to Module Overview Page (https://onlinelearning.berkeley.edu/courses/665040/pages/module-nine)

9.1: Prologue

The most common format used to exchange information over the Internet is HTML (Hypertext Markup Language). As a flexible and easy to use markup language, it quickly became the standard and the preferred format.

HTML is fine for a person reading on a web browser, but is unsuitable for exchanging data between applications.

Imagine if you need to send and receive specific sets of data (a DVD collection, for example). How would you (or your client application) know which part is the title, the performer's name and the price? How is the data delimited? We would need a way to tag and define the data.

This is where XML comes in. XML stands for eXtensible Markup Language. XML is a markup language just like HTML, in that the data is annotated with tags. Tags are metadata that describe a piece of data. In HTML, tags help a web browser determine how to display a piece of text. In XML, it can be whatever the programmer defines it to be.

In this module, we will talk about the XML document and its parts. We will then talk about the Document Type Definition (DTD), which describes the specifications of XML documents. And finally, we will discuss the Java APIs for handling XML documents. Java provides several APIs for creating and parsing XML document. We will examine two very different APIs in detail: DOM (Document Object Model) and StAX (Streaming API for XML).