What are the implications of the concept of the document tree?

Author: Mihai Rotaru

Date: 1 Dec 2011

The concept of document tree is enforced in XML files. Most of the times, it can also be correctly extracted from HTML files, where it is usually accessed and manipulated via JavaScript using the Document Object Model (DOM) API. The term 'document tree' is usually used in the context of XML files, while the DOM is associated mostly with HTML files - although the two terms can, and are sometimes used interchangeably.

In the context of XML files, the concept of the document tree is crucial to a number of aspects. First of all, every XML file contains (or can be considered as being) a document tree - with the root element as the root of the document tree, and each subsequent element forming a branch. The last element on a branch is called a 'leaf'.

One of the implications concerns XML parsers and accessing API's - due to it's tree structure, XML documents are commonly parsed and stored into memory as a tree data structure, with each node representing an element. The DOM is the most common convention for representing and interacting with XML documents.

The document tree is also relied on by XPath - which is used to address parts of an XML document. XPath uses a path notation (hence the name) similar to URL notations to navigate through the hierarchical structure of an XML document - the document tree.

XSL relies on XPath - and therefore, on the document tree - to select elements which to style and modify.