

XML – Usando DOM (Document Object Model) como processador

Os processadores XML são programas de software usados para analisar documentos XML e fornecem acesso ao conteúdo e à estrutura do documento. Os processadores XML também podem validar a estrutura de documentos XML e transformá-la em vários formatos.

Os documentos XML são arquivos de texto que contêm dados formatados como uma série de elementos e atributos. Os elementos em um documento XML também podem ser considerados nós do documento. O elemento raiz do documento é o nó raiz e todos os outros elementos são nós filhos (ou descendentes) desse nó raiz.

Os atributos são dados adicionais associados a um elemento e são considerados propriedades do elemento. Quando um processador XML carrega um documento XML pela primeira vez, ele analisa os nós do documento e os coloca na memória de acesso aleatório (RAM) na forma de um DOM.

Depois que um documento XML tiver sido analisado e carregado no DOM, você poderá usar as interfaces através de programação do DOM para acessar e manipular o documento XML que ele contém. Se uma XSLT (Transformação em XSL) for associada a um documento XML, o processador XML poderá ser usado para aplicá-la ao documento e produzir o formato de saída especificado.

Pense em DOM como se fosse um conjunto de objetos e métodos específicos para manipulação e acesso de um arquivo estruturado em tags, como um documento HTML ou XML.

Origem de DOM

DOM foi inspirado em DHTML, que é uma combinação de HTML, folha de estilo e uma linguagem de script, que juntas permitem a criação de documentos animados.

Esta funcionalidade de DHTML foi o fator que motivou a criação de DOM, porém não confunda DOM com DHTML, pois DOM é uma interface independente de linguagem e plataforma que permite programas e scripts acessarem dinamicamente o conteúdo, estrutura e estilo de um documento XML ou HTML.

DOM é uma representação de um documento XML em uma estrutura em árvore carregada na memória. Portanto as definições das interfaces DOM descrevem os meios de acessarmos esta estrutura em árvore.

Como DOM é Estruturado

DOM é separado em três partes: Core, XML e HTML.

Módulo Core DOM

É formada por um conjunto de objetos que representam um documento XML ou HTML. Esta interface é bem compacta, projetada com estruturas básicas de manipulação do conteúdo de um documento.

Módulo XML e Módulo HTML

Os módulos XML e HTML são formados por interfaces de alto nível que abstraem objetos e métodos, que utilizados junto com o módulo Core DOM permitem uma forma trivial de acesso e manipulação de um documento.

DOM foi projetado em níveis, de forma que à medida que as tecnologias evoluem é permitido que novas características sejam embutidas nas definições de DOM. É importante em conhecer a forma como DOM está estruturado, pois assim você tem consciência do que pode ser feito e o que não pode ser feito com DOM até então.

Abaixo estão descritas estas estruturas e suas respectivas funções:

Nível 1: Este nível engloba os módulos Core, HTML e XML. Conseqüentemente neste nível estão definidas as funcionalidades de navegação e manipulação de documentos.

Nível 2: Inclui modelo de objetos que define funcionalidades para manipulação dos estilo de formatação de documentos. Neste nível também são definidas funções que permitem navegar através de um documento, tratar eventos em documentos e dão suporte a espaços de nome (namespace) XML.

Nível 3: Inclui funcionalidades para salvar e carregar documentos, além de prover suporte a validação de documentos. Neste nível também são definidas as funcionalidades de manipulação de eventos sobre um documento.

Note que nem todos os parsers XML que implementam DOM possuem suporte aos três níveis acima. Ao tomar a decisão em utilizar DOM em seu projeto deve-se observar no manual de seus parsers XML quais níveis estão implementados, pois cada um possui além das características definidas pela especificação de DOM, suas próprias características, e isto pode influenciar substancialmente no desenvolvimento do seu projeto.