
Mòdul 6. Accés a dades

UF1. Persistència en fitxers

Analitzadors sintàctics XML

Anàlisi XML

L'anàlisi XML o *parsing* és el procés que emmarca el processament d'un document XML per extreure informació estructurada i poder utilitzar-la.

Podem dir que el *parsing XML* és el procés en el que s'analitza un document XML per a comprovar si el seu format és correcte, i posteriorment tractar el seu contingut per a ser utilitzat.

Per tant permet a les aplicacions poder llegir, interpretar i utilitzar dades emmagatzemades en documents XML.

Parsing XML

- Lectura del document XML.
 - Anàlisi sintàctic.
 - Construcció de l'estructura interna.
 - Accés i manipulació de dades.
 - Utilització de les dades.
-

Parsing XML

Lectura del document XML.

L'analitzador o parser XML llegeix el document XML d'entrada.

Anàlisi sintàctic.

L'analitzador verifica la sintaxi del document XML per assegurar que compleix l'estructura correcta.

Parsing XML

Construcció de l'estructura interna.

Un cop validada la sintaxi, l'analitzador construeix l'estructura interna de dades que representa el contingut del document XML llegit.

Accés i manipulació de les dades.

Un cop tenim l'estructura interna construïda es pot accedir i manipular les dades que conté el document XML. Lectura d'atributs, cerca de dades, modificació, ...

Parsing XML

Utilització de les dades.

L'aplicació utilitza les dades extretes del document XML per a les operacions o funcionalitats que requereixi.

Mostrar, emmagatzemar, processar, etc ...

Processament fitxer XML

- DOM (Document Object Model)
 - SAX (Simple API for XML)
-

Processament fitxer XML

DOM (Document Object Model)

Facilita una representació estructurada del document i defineix la manera en què els programes hi poden accedir, per tal de modificar la seva estructura i el seu contingut.

DOM dóna una representació del document com un grup de nodes i objectes estructurats que té unes propietats i mètodes.

Processament fitxer XML

DOM (Document Object Model)

Carrega tot el document XML a memòria en forma d'arbre jeràrquic. Per tant el document al complet és accessible i manipulable com a objectes.

Proporciona una visió completa del document i facilita la seva navegació i manipulació.

És poc eficient en termes de memòria i rendiment quan tractem documents XML molt grans.

Processament fitxer XML

SAX (Simple API for XML)

Originalment era una API només per al llenguatge JAVA.

Actualment hi ha versions per a altres llenguatges de desenvolupament.

S'encarrega de recórrer l'estructura del document generant esdeveniments que corresponguin als elements que va trobant.

Processament fitxer XML

SAX (Simple API for XML)

L'anàlisi està basada en esdeveniments. Recorre el document de manera seqüencial i notifica quan troba un element.

És eficient en termes de memòria ja que no carrega tot el document XML, per tant és útil quan s'han de processar documents XML molt grans.

No permet la manipulació fàcil de tot el contingut, ja que es tracta la informació a mesura que avança en el recorregut del document.

Diferències principals entres SAX i DOM

	SAX	DOM
Mode de processament	Basat en esdeveniments. Processament seqüencial.	Carrega el fitxer a memòria. Transformació a arbre d'elements.
Ús de memòria	No consumeix tanta memòria. Processa XML sense la necessitat de carregar-lo a memòria.	Consumeix més memòria ja que carregat tot el document XML abans de processar-lo.
Accés al document	Només accés seqüencial. Només lectura. No es pot accedir aleatòriament als elements.	Permet l'accés aleatori a qualsevol part del document. Facilita la seva modificació ja que et permet modificar i afegir o eliminar nodes.
Velocitat	Generalment més ràpid al no carregar el fitxer a memòria. Però depèn del tamany del fitxer i el tipus de tasca.	Més lent en càrrega inicial del document, però més eficient en accés a múltiples parts del document.
Manipulació del document	No permet modificar el document.	Permet modificar, un cop està carregat a memòria.
Complexitat	Potser més complex d'implementar, ja que s'ha de gestionar els esdeveniments manualment.	Fàcil d'utilitzar per accedir o modificar el document, gràcies a la seva estructura d'arbre.
Utilització	Útil quan treballes amb documents molt grans i només tens la necessitat de llegir-los.	Útil quan necessites accedir en parts concretes del document per a fer certes accions.