



M4 Llenguatges de marques i sistemes de gestió d'informació

UF2-NF2.1. CONVERSIÓ DE DOCUMENTS XML

Continguts

1. Introducció
 - 1.2. Eines per la conversió (XSL-FO)
2. XPATH
3. XSLT
4. Elaboració de documentació tècnica

1. Introducció:

XML està pensat sobretot per a **emmagatzemar i intercanviar informació**, de manera que si cal representar les dades d'una manera diferent per optimitzar un procés o per millorar-ne la **visualització** hi haurà **diverses possibilitats**:

- **Desenvolupar un programa:** es podria desenvolupar un programa que agafi les dades XML i generi la sortida tal com la volem. Això té l'inconvenient que caldrà tenir **coneixements de programació** i que pot representar molta feina.
- **Fer servir CSS:** representar la informació de manera més amigable fent servir un navegador. Però només serveix per **canviar la visualització**, **no** per canviar el **document**.
- **Transformar el document:** transformar el document en un altre que estigui pensat per ser visualitzat. Hi ha molt formats que estan pensats sobretot per ser visualitzats: **PDF, HTML, XHTML, etc.**

1.2. Eines per la conversió:

Per intentar aconseguir fer tot allò que CSS no podia fer es va crear un nou **llenguatge de plantilles: XSL** (extensible stylesheet language).

Aquest permet **transformacions genèriques de documents XML** en altres coses: documents de text, documents XML, pàgines web, etc.

XSL

És una família de llenguatges que serveixen per definir transformacions i presentacions de documents XML.

La família **XSL** està formada per **tres llenguatges**:

- **XPATH:** un llenguatge per accedir a parts dels documents XML.
- **XSLT** (*XSL transformations*): un llenguatge per transformar documents XML.
- **XSL-FO** (XSL formatting objects): un llenguatge per definir el format que s'ha d'aplicar a un document. (documentació tècnica).

XSL-FO:

XSL-FO és un llenguatge basat en XML que està pensat per **donar format als documents XML**, però a diferència d'altres llenguatges amb objectius similars, com CSS o XHTML, està pensat per generar **sortides tant per a formats de pantalla com per a formats paginats**.

XSL-FO és un llenguatge que:

- Permet especificar amb molta precisió el contingut d'un document (paginació, organització, estil, etc.).
- Permet crear documents d'alta qualitat.
- És ideal per generar documents amb dades que canvien sovint.

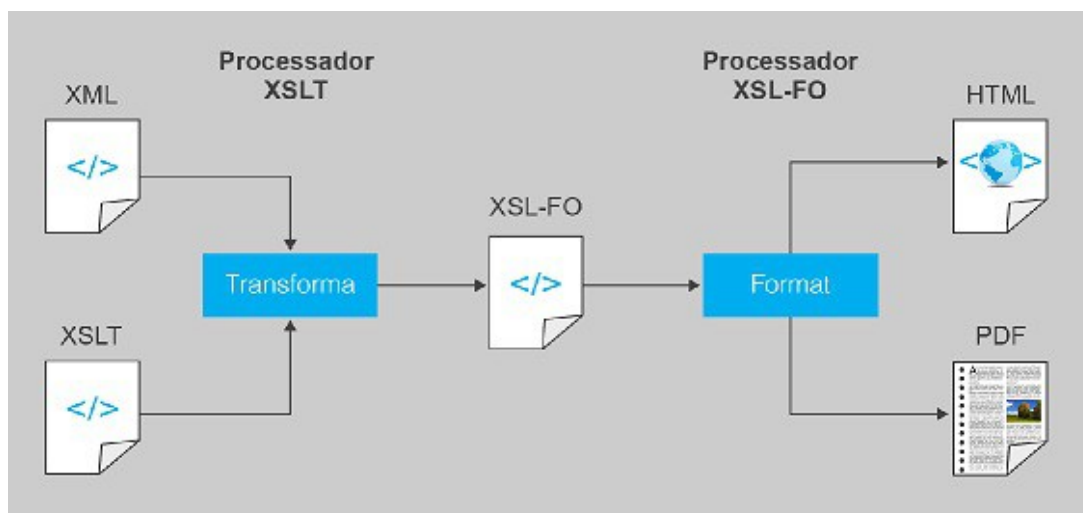
XSL-FO sobretot es fa servir per **generar documents** en formats pensats **per ser impresos** com **PDF o Postscript**.

Apache FOP

És un dels **processadors** d'XSL-FO més populars, ens permet crear documents en PDF a partir de documents XSL-FO.

La generació del document es fa en dues fases:

1. **Transformació del document XML en un document XSL-FO** amb el llenguatge de transformacions XSLT.
2. **Transformació del document XSL-FO en el format que volem** amb un processador XSL-FO.



2. XPATH:

XPath (XML *path language*)

És una manera d'especificar parts d'un document XML que té eines per manipular el contingut de les dades de text, numèriques, etc.

La base del funcionament d'XPath és l'**avaluació d'expressions**. Una expressió que s'avaluarà contra **un document XML** i ens donarà un resultat que pot ser de diferents tipus:

1. Un booleà: cert o fals
2. Un nombre
3. Una cadena de caràcters
4. Un grup d'elements

XPath està desenvolupat pels comitès de creació d'XSL i XQuery i s'ha convertit en un component essencial per a diferents llenguatges XML com XLinks, XSLT i Xquery.

La versió 2.0 d'XPath està tan integrada dins d'XQuery que qualsevol **expressió XPath** és també automàticament una **expressió XQuery correcta**.

<http://www.w3.org/TR/xpath>

➤ Vista Arbre:

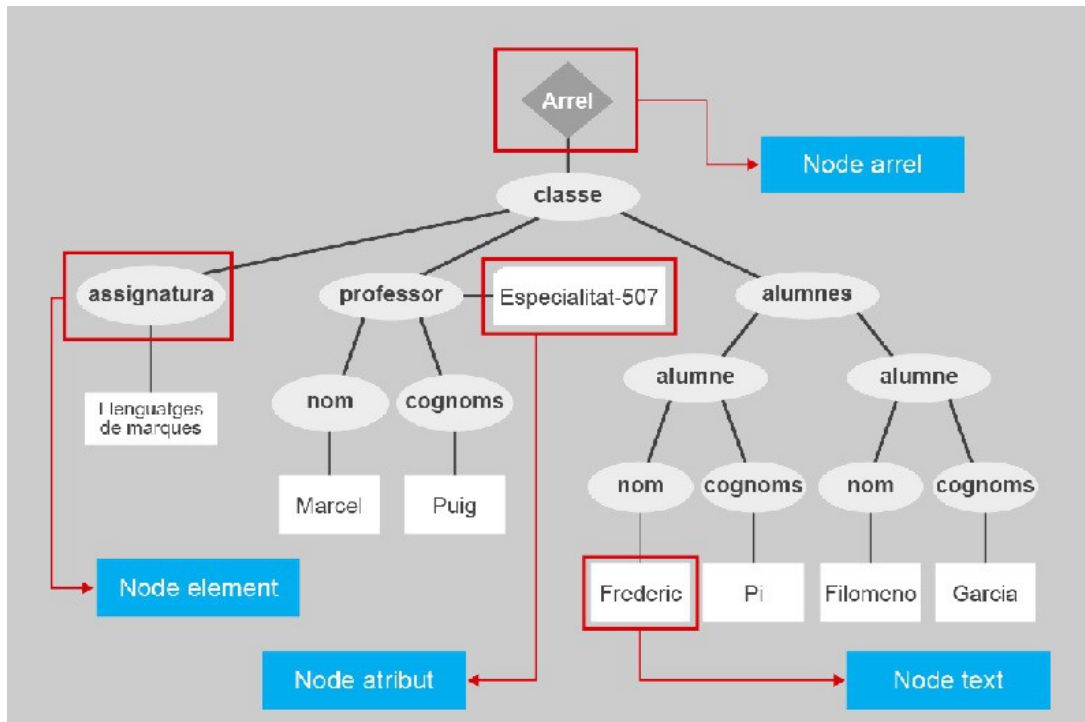
XPath tracta tots els documents XML des del punt de vista d'un arbre de nodes en què hi ha una **arrel que no es correspon amb** l'arrel del document, sinó que és **el document** i es **representa amb el símbol “/”**.

A part de l'arrel també hi ha **nodes** per representar els elements, els atributs, els nodes de dades, els comentaris, les instruccions de procés i els espais de noms.

Exemple XML:

```
<?xml version="1.0" ?>
<classe>
  <assignatura>Llenguatges de marques</assignatura>
  <professor Especialitat="507">
    <nom>Marcel</nom>
    <cognoms>Puig</cognoms>
  </professor>
  <alumnes>
    <alumne>
      <nom>Frederic</nom>
      <cognoms>Pi</cognoms>
    </alumne>
    <alumne>
      <nom>Filomeno</nom>
      <cognoms>Garcia</cognoms>
    </alumne>
  </alumnes>
</classe>
```

Es representarà en XPath amb un arbre com el següent:



En un arbre XPath els **atributs** no són considerats nodes fills sinó que són **“propietats” del node** que els conté i els nodes de dades són nodes sense nom que només contenen les dades.

➤ Avaluat XPath:

Molts programes permeten executar una consulta XPath contra un document XML. Hi ha programes específics, editors XML, components dels navegadors web, pàgines web online, etc.

Programes que permeten treballar amb Xpath:

- **"xpath"**: La instal·lació per defecte d'**Ubuntu instal·la automàticament** una ordre anomenada xpath que permet comprovar les ordres XPath des de la **consola**. L'expressió es posa rere el paràmetre -e.
- **XPath Checker**: És un **component del Firefox**, que permet fer consultes XPath interactivament, tant en pàgines HTML com en documents XML que s'hagin visualitzat en el navegador.
- **Editors XML**: opció més professional. Aquests editors **permeten avaluar** expressions XPath **des d'un entorn gràfic** amb tot tipus d'assistències en la creació de les expressions, possibilitats de depuració, etc. Exemple: l'oXygen XML Editor.

➤ Navegació XPATH:

Expressar camins en XPath s'assembla tant a com ho fan els sistemes operatius que fins i tot es fan servir símbols semblants.

El més important per tenir en compte a l'hora de crear una expressió XPath és saber el node en el qual està situat el procés (**node de context**)

Podem expressar els camins XPath de dues maneres:

- **Camins absoluts:** parteixen de l'arrel de l'arbre. Es poden identificar perquè el primer caràcter de l'expressió sempre serà **"/"**. No importa quin sigui el node de context si es fan servir camins absoluts, perquè el resultat sempre serà el mateix.

- **Camins relatius:** parteixen des del node en el qual estem situats.

➤ Tipus d'expressions i significats: explicarem en base l'exemple anterior.

- **Obtenir un valor del resultat d'una expressió.** Exemple: Obtenir el node `<nom>` del professor.

| Expressió Xpath | Resultat |
|------------------------------------|--|
| <code>/classe/professor/nom</code> | <code><nom>Marcel</nom></code> |

- **Obtenir més d'un valor d'una expressió.**

| Expressió Xpath | Resultat |
|---|--|
| <code>/classe/alumnes/alumne/nom</code> | <code><nom>Frederic</nom></code> <code><nom>Filomeno</nom></code> |

- **Retornar qualsevol tipus d'element com a resultat.** (no cal siguin finals)

| Expressió Xpath | Resultat |
|------------------------------|---|
| <code>/classe/alumnes</code> | <code><alumnes></code> <code><alumne></code> <code><nom>Frederic</nom></code> <code><cognom>Pi</cognom></code> <code></alumne></code> <code><alumne></code> <code><nom>Filomeno</nom></code> <code><cognom>Garcia</cognom></code> <code></alumne></code> <code></alumnes></code> |

- **Seleccionar un element d'un conjunt**

| Expressió Xpath | Resultat |
|---------------------------|---|
| /classe/alumnes/alumne[1] | <alumne> <nom>Frederic</nom> <cognom>Pi</cognom> </alumne> |

- **Determinar quina de les branques es triarà.** Exemple: obtenir només el nom del segon alumne

| Expressió Xpath |
|-------------------------------|
| /classe/alumnes/alumne[2]/nom |

- **Obtenir els atributs d'un element**

| Expressió Xpath | Resultat |
|---------------------------------|----------|
| /classe/professor/@especialitat | 507 |

S'ha de tenir en compte que a diferència del que passa amb els elements, en obtenir un atribut no tindrem un element sinó només el seu valor.

- **Obtenir el contingut d'un element (sense etiquetes)**

| Expressió Xpath | Resultat |
|------------------------------|----------|
| /classe/professor/nom/text() | Marcel |

- **Comodins**

| Comodí | Significat |
|--------|---|
| * | L'asterisc es fa servir per indicar tots els elements d'un determinat nivell. |
| . | Com en els directoris dels sistemes operatius el punt serveix per indicar el node actual. |
| .. | Es fa servir per indicar el pare del node en el qual estem. |
| // | Les dobles barres indiquen que quadrarà amb qualsevol cosa des del node en el qual estem. Pot ser un sol element o un arbre de nodes. |

| Expressió Xpath | Resultat |
|----------------------|---|
| /classe/professor/* | <nom>Marcel</nom> <cognoms>Puig</cognoms> |
| //nom | <nom>Marcel</nom> <nom>Frederic</nom> <nom>Filomeno</nom> |
| /classe/alumnes//nom | <nom>Frederic</nom> <nom>Filomeno</nom> |

- **Condicions:**

| Expressió Xpath | Significat |
|--|--|
| /classe/professor[../alumnes/alumne] | Els professors que tenen alumnes. |
| /classe/professor[nom="Marcel"] | Els professors que es diguin “Marcel” |
| /classe/professor[nom="Marcel"]/cognoms | El cognom del professor que es diu “Marcel” |
| /classe/professor[@especialitat=507] | El professor que té una especialitat determinada |
| /classe/professor[@especialitat="507"] [nom="Marcel"] | El professor que té l'atribut especialitat a “507” i que es diu “Marcel” |

- **Seqüències:** s’avaluen en ordre, primer la primera expressió, després la segona, etc.

| Expressió Xpath | Resultat |
|----------------------------------|--|
| (//nom/text(), //cognoms/text()) | Marcel Frederic Filomeno Puig Pi Garcia |

- **Unió, intersecció i disjunció:** També es pot operar amb les seqüències d’elements. Una manera seria fer servir els operadors d’unió (**union**), intersecció (**intersect**) o disjunció (**except**).

Exemple: (//alumne/nom) intersect (//professor/nom)

- **Funcions:**

| Funció | Ús |
|-------------------|--|
| name() | Retorna el nom del node |
| sum() | Retorna la suma d'una seqüència de nodes o de valors |
| count() | Retorna el nombre de nodes que hi ha en una seqüència |
| avg() | Fa la mitjana dels valors de la seqüència |
| max() | Retorna el valor màxim de la seqüència |
| min() | Dóna el valor mínim de la seqüència |
| position() | Diu en quina posició es troba el node actual |
| last() | Retorna si el node actual és l'últim |
| distinct-values() | Retorna els elements de la seqüència sense duplicats |
| concat() | Uneix dues cadenes de caràcters |
| starts-with() | Retorna si la cadena comença amb els caràcters marcats |
| contains() | Ens diu si el resultat conté el valor |
| string-length() | Retorna la llargada de la cadena |
| substring() | Permet extreure una subcadena del resultat |
| string-join() | Uneix la seqüència amb el separador especificat |
| current-date() | Ens retornarà l'hora actual |
| not() | Inverteix el valors booleans |

Exemple: “quants alumnes tenim?”

| Expressió Xpath | Resultat |
|-------------------------------|-----------------|
| count(/classe/alumnes/alumne) | 2 |