



Pràctica 8: XPath

Objectius

Els objectius d'aquesta pràctica d'XPath se centren a desenvolupar una comprensió fonamental de la sintaxi i les funcionalitats bàsiques d'XPath. En primer lloc, caldrà familiaritzar-se amb la sintaxi bàsica de l'XPath i adquirir habilitats per seleccionar elements específics d'un document XML mitjançant camins absoluts i relatius. Un segon objectiu important és la pràctica en la manipulació de text i atributs, que inclourà l'extracció de contingut de text d'elements i l'accés als valors dels atributs.

Lliuraments

Els resultats d'aquesta part de la pràctica s'hauran d'entregar en format PDF i l'entrega pot ser a través de GIT* o el moodle.

* S'ha d'entregar l'enllaç del GIT al moodle.

Guió

Exercici 1

Completa la taula explicant els resultats esperats per a cada una de les expressions XPath donades. En casos on la resposta sigui **un objecte sigui substancialment gran**, només cal que indiquis de quin element/s es tracta.

- Si dona **error** o no dona **cap resultat**, explica'n la raó

Exercici 2

Una vegada tinguis la taula de l'exercici 1 plena, avalua les expressions XPath fent servir el codi Python 3 que es troba a l'arxiu *xpath_evaluator.py*.

- Assegura't tenir instal·lada la llibreria 'lxml'. Per instal·lar-la, pots executar al terminal: `pip install lxml`

	Ruta XPath	Explica el resultat
1	/llenguatges/llenguatge/nom	<i>No da error. Elemento nom dentro de lenguaje</i>
2	/llenguatges/llenguatge/nom/node()	<i>No da error. Todos los nodos nom dentro de lenguajes</i>
3	/llenguatges/llenguatge/paradigmes[@tipat]	<i>No da error. Todos los paradigmes con atributo tipat</i>
4	/llenguatges/llenguatge/paradigmes[@tipat="false"]	<i>Error expresion invalida</i>
5	nom	<i>No da error. El contenido del elemento nom</i>
6	/nom	<i>No da error. El contenido del elemento nom en la raíz</i>
7	/nom/	
8	//nom	<i>No da error. Todos los elementos nom del documento</i>
9	//nom/node()	<i>No da error. Todos los nodos nom del documento</i>
10	//llenguatge/nom	<i>No da error. Todos los nom de lenguaje</i>
11	//llenguatge/nom/node()	<i>No da error. Todos los nodos de nom de lenguaje</i>
12	//nom/node() //popularitat/node()	
13	//mode_execucio/*	
14	//llenguatge[2]	<i>No da error. El segundo elemento lenguaje del documento</i>
15	//llenguatge[last()-1]	<i>No da error. El ultimo elemento, menos 1 de lenguaje en el documento</i>
16	//llenguatge[@fundacio]	<i>No da error. Todos los lenguajes que tengan fundacio como atributo</i>
17	//@fundacio	<i>No da error. Todos los elementos con atributo de fundacio del documento</i>

18	//@fundacio[.>2000]	<i>No da error. Todos los elementos con atributo de fundacio del documento con un valor superior a 2000</i>
19	//@fundacio[.>2000]/nom	<i>No da error. Todos los elementos nom con atributo de fundacio del documento con un valor superior a 2000</i>
20	//@fundacio[.>2000]/../nom	<i>No da error. Todos los elementos con atributo de fundacio del documento con un valor superior a 2000 con elemento padre nom</i>
21	//llenguatge[mode_execucio='Java Virtual Machine'] /nom/text()	
22	//llenguatge[nom='Kotlin']/popularitat/node()	
23	//mode_execucio[.='Interpretat']/..	
24	//paradigmes[node()='Imperatiu']/../nom	<i>No da error.</i>
25	//*	<i>No da error. Todos los elementos</i>

```
<?xml version="1.0" encoding="UTF-8" ?>
<llenguatges>
  <llenguatge fundacio="2011">
    <nom>Kotlin</nom>
    <mode_execucio>Java Virtual Machine</mode_execucio>
    <popularitat>Creixent</popularitat>
    <paradigmes tipat="true">
      <paradigma>Orientat a objectes</paradigma>
      <paradigma>Imperatiu</paradigma>
      <paradigma>Funcional</paradigma>
    </paradigmes>
  </llenguatge>
  <llenguatge fundacio="1983">
    <nom>C++</nom>
    <mode_execucio>Compilació</mode_execucio>
    <popularitat>Alta</popularitat>
    <paradigmes tipat="true">
      <paradigma>Procedimental</paradigma>
      <paradigma>Imperatiu</paradigma>
      <paradigma>Orientat a objectes</paradigma>
      <paradigma>Programació genèrica</paradigma>
    </paradigmes>
  </llenguatge>
</llenguatges>
```

```

<llenguatge fundacio="1995">
  <nom>Java</nom>
  <mode_execucio>Java Virtual Machine</mode_execucio>
  <popularitat>Alta</popularitat>
  <paradigmes tipat="true">
    <paradigma>Orientat a objectes</paradigma>
    <paradigma>Basat en classes</paradigma>
  </paradigmes>
</llenguatge>
<llenguatge fundacio="1991">
  <nom>Python</nom>
  <mode_execucio>Interpretat</mode_execucio>
  <popularitat>Molt alta</popularitat>
  <paradigmes tipat="false">
    <paradigma>Imperatiu</paradigma>
    <paradigma>Funcional</paradigma>
  </paradigmes>
</llenguatge>
<llenguatge fundacio="1995">
  <nom>JavaScript</nom>
  <mode_execucio>Interpretat</mode_execucio>
  <popularitat>Alta</popularitat>
  <paradigmes tipat="false">
    <paradigma>Orientat a objectes</paradigma>
    <paradigma>Funcional</paradigma>
  </paradigmes>
</llenguatge>
<nom>Això no és un llenguatge de programació :(</nom>
</llenguatges>

```

Después de ejecutar el script y ajustar la ruta xpath de cada uno a la tabla, los resultados a la derecha

1	/llenguatges/llenguatge/nom	<nom>C++</nom> <nom>Java</nom> <nom>Python</nom> <nom>JavaScript</nom>
2	/llenguatges/llenguatge/nom/node()	Kotlin C++ Java Python JavaScript
3	/llenguatges/llenguatge/paradigmes[@tipat]	<paradigmes tipat="true">

		<p><paradigma>Orientat a objectes</paradigma></p> <p><paradigma>Imperatiu</paradigma></p> <p><paradigma>Funcional</paradigma></p> <p></paradigmes></p> <p><paradigmes tipat="true"></p> <p><paradigma>Procedimental</paradigma></p> <p><paradigma>Imperatiu</paradigma></p> <p><paradigma>Orientat a objectes</paradigma></p> <p><paradigma>Programaci&#243; gen&#232;rica</paradigma></p> <p></paradigmes></p> <p><paradigmes tipat="true"></p> <p><paradigma>Orientat a objectes</paradigma></p> <p><paradigma>Basat en classes</paradigma></p> <p></paradigmes></p> <p><paradigmes tipat="false"></p> <p><paradigma>Imperatiu</paradigma></p> <p><paradigma>Funcional</paradigma></p> <p></paradigmes></p> <p><paradigmes tipat="false"></p> <p><paradigma>Orientat a objectes</paradigma></p> <p><paradigma>Funcional</paradigma></p> <p></paradigmes></p>
4	/llenguatges/llenguatge/paradigmes[@tipat="false"]	error
5	nom	<p><nom>Aix&#242; no &#233;s un llenguatge de programaci&#243; :(</nom></p>

6	/nom	ningun resultado
7	/nom/	Error
8	//nom	<nom>Kotlin</nom> <nom>C++</nom> <nom>Java</nom> <nom>Python</nom> <nom>JavaScript</nom> <nom>Això no és un llenguatge de programació :(</nom>
9	//nom/node()	Kotlin C++ Java Python JavaScript Això no és un llenguatge de programació :(
10	//llenguatge/nom	<nom>Kotlin</nom> <nom>C++</nom> <nom>Java</nom> <nom>Python</nom> <nom>JavaScript</nom>
11	//llenguatge/nom/node()	Kotlin C++ Java Python JavaScript
12	//nom/node() //popularitat/node()	Kotlin Creixent C++ Alta Java Alta Python Molt alta

		<i>JavaScript</i> <i>Alta</i> <i>Això no és un llenguatge de programació :(</i>
13	//mode_execucio/*	ningun resultat
14	//llenguatge[2]	<llenguatge fundacio="1983"> <nom>C++</nom> <mode_execucio>Compilació</mode_execucio> <popularitat>Alta</popularitat> <paradigmes tipat="true"> <paradigma>Procedimental</paradigma> <paradigma>Imperatiu</paradigma> <paradigma>Orientat a objectes</paradigma> <paradigma>Programació genèrica</paradigma> </paradigmes> </llenguatge>
15	//llenguatge[last()-1]	<llenguatge fundacio="1991"> <nom>Python</nom> <mode_execucio>Interpretat</mode_execucio> <popularitat>Molt alta</popularitat> <paradigmes tipat="false"> <paradigma>Imperatiu</paradigma> <paradigma>Funcional</paradigma> </paradigmes> </llenguatge>
16	//llenguatge[@fundacio]	<llenguatge fundacio="2011"> <nom>Kotlin</nom> <mode_execucio>Java

		<p>Virtual Machine</mode_execucio></p> <p><popularitat>Creixent</popula ritat></p> <p> <paradigmes tipat="true"></p> <p> <paradigma>Orientat a objectes</paradigma></p> <p><paradigma>Imperatiu</parad igma></p> <p><paradigma>Funcional</parad igma></p> <p> </paradigmes></p> <p></llenguatge></p> <p><llenguatge fundacio="1983"></p> <p> <nom>C++</nom></p> <p><mode_execucio>Compilaci&#2 43;</mode_execucio></p> <p><popularitat>Alta</popularitat ></p> <p> <paradigmes tipat="true"></p> <p><paradigma>Procedimental</p aradigma></p> <p><paradigma>Imperatiu</parad igma></p> <p> <paradigma>Orientat a objectes</paradigma></p> <p><paradigma>Programaci&#243 ; gen&#232;rica</paradigma></p> <p> </paradigmes></p> <p></llenguatge></p> <p><llenguatge fundacio="1995"></p> <p> <nom>Java</nom></p> <p> <mode_execucio>Java Virtual Machine</mode_execucio></p> <p><popularitat>Alta</popularitat ></p> <p> <paradigmes tipat="true"></p> <p> <paradigma>Orientat a objectes</paradigma></p> <p> <paradigma>Basat en classes</paradigma></p> <p> </paradigmes></p>
--	--	---

		</llenguatge> <llenguatge fundacio="1991"> <nom>Python</nom> <mode_execucio>Interpretat</mode_execucio> <popularitat>Molt alta</popularitat> <paradigmes tipat="false"> <paradigma>Imperatiu</paradigma> <paradigma>Funcional</paradigma> </paradigmes> </llenguatge> <llenguatge fundacio="1995"> <nom>JavaScript</nom> <mode_execucio>Interpretat</mode_execucio> <popularitat>Alta</popularitat> <paradigmes tipat="false"> <paradigma>Orientat a objectes</paradigma> <paradigma>Funcional</paradigma> </paradigmes> </llenguatge>
17	//@fundacio	2011 1983 1995 1991 1995
18	//@fundacio[.>2000]	2011
19	//@fundacio[.>2000]/nom	ningun resultado
20	//@fundacio[.>2000]/../nom	<nom>Kotlin</nom>
21	//llenguatge[mode_execucio='Java Virtual Machine']/nom/text()	error

22	<code>//llenguatge[nom='Kotlin']/popularitat/node()</code>	<i>Error</i>
23	<code>//mode_execucio[.='Interpretat']/..</code>	<i>error</i>
24	<code>//paradigmes[node()='Imperatiu']/../nom</code>	<i>error</i>
25	<code>//*</code>	<i>todo el documento que no voy a pegar 😊</i>

Exercici 3

El fitxer **ods.xml** conté els **Objectius de Desenvolupament Sostenible (ODS)** en català. Els ODS són un conjunt de 17 objectius interconnectats adoptats per les Nacions Unides per abordar els reptes mundials, com la pobresa, la desigualtat, el canvi climàtic i la justícia social, amb l'objectiu de millorar la vida de les persones i protegir el planeta.

<https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

Dissenyeu l'expressió XPath per accedir a la informació sol·licitada al fitxer *ods.xml*. Recordeu que heu de satisfer les condicions de l'enunciat i, tot i conèixer l'XML, heu d'extreure la informació sense afegir cap informació addicional. Per exemple, si se us demana seleccionar l'ODS titulat "Igualtat de Gènere", **heu de seleccionar-lo exclusivament fent servir aquesta informació**. L'exercici seria incorrecte si seleccioneu la informació fent servir el fet que és l'ods=5.

1. El text de dins de l'etiqueta **<titol>** de **tots els ODS** (Objectius de Desenvolupament Sostenible).

Fi de la Pobresa: `//objectiu[titol="Fi de la Pobresa"]/titol/text()`

Fam Zero: `//objectiu[titol="Fam Zero"]/titol/text()`

Salut i Benestar: `//objectiu[titol="Salut i Benestar"]/titol/text()`

Educació de Qualitat: `//objectiu[titol="Educació de Qualitat"]/titol/text()`

Igualtat de Gènere: `//objectiu[titol="Igualtat de Gènere"]/titol/text()`

Aigua Neta i Sanejament: `//objectiu[titol="Aigua Neta i Sanejament"]/titol/text()`

Reducció de les Desigualtats: `//objectiu[titol="Reducció de les Desigualtats"]/titol/text()`

Ciutats i Comunitats Sostenibles: `//objectiu[titol="Ciutats i Comunitats Sostenibles"]/titol/text()`

Pau, Justícia i Institucions Sòlides: //objectiu[titol="Pau, Justícia i Institucions Sòlides"]/titol/text()
Energia Assequible i No Contaminant: //objectiu[titol="Energia Assequible i No Contaminant"]/titol/text()
Indústria, Innovació i Infraestructura: //objectiu[titol="Indústria, Innovació i Infraestructura"]/titol/text()
Producció i Consum Responsables: //objectiu[titol="Producció i Consum Responsables"]/titol/text()
Acció pel Clima: //objectiu[titol="Acció pel Clima"]/titol/text()
Vida Submarina: //objectiu[titol="Vida Submarina"]/titol/text()
Vida d'Ecosistemes Terrestres: //objectiu[titol="Vida d'Ecosistemes Terrestres"]/titol/text()
Treball Digne i Creixement Econòmic: //objectiu[titol="Treball Digne i Creixement Econòmic"]/titol/text()
Aliances per a Assolir els Objectius: //objectiu[titol="Aliances per a Assolir els Objectius"]/titol/text()

2. La descripció (text) de l'ODS titulat “Igualtat de Gènere”.

Promou la igualtat de gènere i empoderar totes les dones i nenes. Inclou objectius com eliminar la violència de gènere i garantir la participació igualitària en la presa de decisions.

//objectiu[titol="Igualtat de Gènere"]/descripcio/text()

3. Les accions (text) de l'ODS que té el títol de “Fi de la Pobresa”.

Implementar polítiques socials inclusives
Garantir protecció social per a tots

//objectiu[titol='Fi de la Pobresa']/accions/text()

4. El títol (text) de les ODS 13, 14, 15 i 16.

Pau, Justícia i Institucions Sòlides
Acció pel Clima
Vida Submarina
Vida d'Ecosistemes Terrestres

//objectiu[numero=13 or numero=14 or numero=15 or numero=16]/titol/text()

5. El **títol** i la **descripció** (text) de l'ODS amb **ods="10"**.

Reducció de les Desigualtats

Busca reduir les bretxes entre països i dins d'ells. Inclou objectius com empoderar les persones marginades i promoure polítiques inclusives.

```
//objectiu[@ods='10']/titol/text() | //objectiu[@ods='10']/descripcio/text()
```

6. El llistat d'**accions**, en XML, dels ODS que pertanyen al grup de **tipus econòmic**.

```
<accio>Fomentar l'emprenedoria i la innovació</accio>
<accio>Reduir la bretxa salarial de gènere</accio>
<accio>Promoure la cooperació internacional en el desenvolupament</accio>
<accio>Facilitar l'accés a la tecnologia i la innovació</accio>
```

```
//objectiu[tipus='econòmic']/accions
```

7. El text de dins de l'etiqueta **<titol>** del **5è objectiu** dins del grup d'ODS de tipus **"ambiental"**.

Vida Submarina

```
//objectiu[tipus='ambiental'][5]/titol/text()
```

8. Dins el grup de tipus **social**, selecciona el **7è objectiu** i retorna la **2a acció**.

```
<accio>Garantir igualtat d'oportunitats per a tothom</accio>
```

```
//objectiu[tipus='social'][7]/accions[2]
```

9. El **tipus del grup** al qual pertany l'ODS que té l'acció **"Promoure l'ús d'energies renovables"**.

ambiental

```
//objectiu[accions='Promoure l'ús d'energies renovables']/tipus
```

10. Busca les etiquetes amagades **<start>** i **<end>** (fes ctrl+F) i troba el camí per començar a l'etiqueta **<start>** i arribar a imprimir el text que hi ha dins l'etiqueta **<end>**. La teva ruta ha de començar per **//start** i has d'intentar fer l'expressió XPath més breu possible.

```
<end>Busca promoure l'ocupació digna, el creixement econòmic inclusiu i la protecció social.
Inclou objectius com reduir la informalitat laboral i garantir salaris justos.</end>
```

11. Les **5 primeres accions** per les ODS de tipus **ambiental**. [opcional]

```
<accio>Desenvolupar infraestructures per a energies renovables</accio>
<accio>Promoure l'eficiència energètica a la indústria</accio>
<accio>Invertir en investigació i desenvolupament</accio>
```

```
<accio>Desenvolupar infraestructures de transport sostenible</accio>  
<accio>Fomentar el reciclatge i la reutilització de productes</accio>
```

```
//objectiu[tipus='ambiental']/accions[position() <= 5]
```

PART 8.2

Exercici 1

Per començar, clona el repositori de GIT que es troba en aquesta ubicació i executa el codi Python per veure quin resultat dona.

https://github.com/pauitic/practica8_2

```
Create Mode 100044 file (505/240c11c8-8 David Ramirez.ppt)
david@david-GP62M-7RD:~/Escritorio/M4_lenguamarcas$ git clone https://github.com
/pauitic/practica8_2
Clonando en 'practica8_2'...
remote: Enumerating objects: 12, done.
remote: Counting objects: 100% (12/12), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 12 (delta 3), reused 12 (delta 3), pack-reused 0
Recibiendo objetos: 100% (12/12), listo.
Resolviendo deltas: 100% (3/3), listo.
```

Exercici 2

- a. Executa les següents rutes XPath i observa el resultat que dona cada una. A continuació, explica les diferències que hi ha entre cada resultat i raona per què produeixen resultats diferents.

- i. node() vs text()

Ruta 1: `//div[@class='attribution']/p/node()`

© 2022

All Rights Reserved.

.

Created with Free
Html Templates.

.

Ruta 2: `//div[@class='attribution']/p/text()`

© 2022

.

.

La diferencia es que text() solo devuelve los nodos de texto que no

están en tags y node si porque selecciona todos los nodos

ii. Barra simple vs barra doble

Ruta 1: `//ul[@class='navbar-nav']/li/a/text()`

Ruta 2: `//ul[@class='navbar-nav']//li/a/text()`

Mientras la ruta 1 espera que busque el texto dentro de los elementos a que son hijos directos de li, en la Ruta 2 es más flexible porque son elementos a descendientes de li.

b. Representa, en forma d'arbre l'estructura XML que resulta d'avaluar la següent ruta XPath (pots ignorar els salts de línia i espais).

i. `(//div/h5) [6]`

```
- h5
  -span
    -text
```

ii. `//div[@class='carousel-item'] [1]//h1`

```
- /
- div
  - [@class='carousel-item']
    - [1]
      - //
        - h1
```

Exercici 3

Descobreix la ruta XPath per arribar a cada un dels elements que es demana tenint en compte només la informació que es proporciona a l'enunciat.

c. Troba la ruta que arriba al **correu** de contacte que es troba al **<footer>** de la pàgina. Comença la ruta a l'etiqueta **<html>**

`/html`

sales@mail.com

`/html/body/footer//p/span[text()='sales@mail.com']`

- d. Troba la ruta que arriba a l'**atribut src** de la següent imatge (n'hi ha una al *<footer>*, i una al *<header>*, pots escollir):



`images/logo.svg`

`/html/body/header//a/img`

- e. Troba la ruta fins a l'**atribut src** de les imatges amb ***alt="Client"***.

`images/client-one.png`

`images/client-two.png`

`images/client-three.png`

`/html/body//img[@alt='Client']/@src`

- f. Troba la ruta fins a l'**adreça** de la pàgina web "***Fake Street 123***". Fes que l'adreça XPath parteixi la següent ubicació:

```
//div[@class='information-f']/p[1]/strong/text()
```

Fake Street 123

- g. Troba la ruta que arriba fins al **<h5>** del "***New Skateboard 12***". **[Pista:** busca la utilitat de la funció *normalize-space()* **]**.

```
<h5>                <span>New Skateboard</span> 12
</h5>
```

`//div[contains(normalize-space(), "New Skateboard 12")]/h5/text()`

- h. Partint de la ruta de l'apartat anterior, Troba la ruta que arriba fins al **preu** (text) del "***New Skateboard 12***".


```
//div[contains(normalize-space(), "New Skateboard 12")]/h6/text()
```

Exercici 4

Canvia la ruta a <https://scrapepark.org/table.html> . Amb l'ajuda del navegador, comprova què hi ha dins d'aquesta pàgina i troba la ruta XPath dels següents elements.

- i. Troba la ruta XPath a tots els **preus dels elements de color 'Blue'**. El resultat ha de ser el següent:

Blue
\$64
\$70
\$80
\$85

```
//tbody/tr[td[1]="Blue"]/td[@class="text-center"]
```

- j. Troba la ruta que imprimeix **els preus del longboard** que es troben a la 4a columna de la taula **pintats en vermell**.

Longboard
\$80
\$85
\$90
\$62
\$150

```
//td[contains(@style, "color: red;")]
```

- k. Indica el nom i color de l'article que val \$110. Comença l'expressió de la següent manera: **[pista]**: hauràs de fer servir l'operador "[]"

```
//td[text()=' $110 ']
```

Skate
Special

```
//th[text()='110$'] | //td[text()='Special'] | //th[text()='Skate']
```

- I. Troba la ruta a **tots els preus** dels objectes "Purple" **excepte el preu** que està pintat en vermell.

```
<td>Purple</td>
```

```
<td class="text-center">$55</td>
```

```
<td class="text-center">$60</td>
```

```
<td class="text-center">$72</td>
```

```
//tbody/tr[td[1]="Purple"]/td[@class="text-center" and  
not(@style="color: red;")]
```