

主旨：

XPath

說明：

XPath 即為 XML 路徑語言（XML Path Language），它是一種用來確定 XML 檔案中某部分位置的電腦語言。XPath 基於 XML 的樹狀結構，提供在資料結構樹中找尋節點的能力。

常用的表示法如下：

Expression	Description
<i>nodename</i>	Selects all nodes with the name " <i>nodename</i> "
/	Selects from the root node
//	Selects nodes in the document from the current node that match the selection no matter where they are
.	Selects the current node
..	Selects the parent of the current node
@	Selects attributes

舉例來說

Path Expression	Result
bookstore	Selects all nodes with the name "bookstore"
/bookstore	Selects the root element bookstore
	Note: If the path starts with a slash (/) it always represents an absolute path to an element!
bookstore/book	Selects all book elements that are children of bookstore
//book	Selects all book elements no matter where they are in the document
bookstore//book	Selects all book elements that are descendant of the bookstore element, no matter where they are under the bookstore element
//@lang	Selects all attributes that are named lang
/bookstore/book[1]	Selects the first book element that is the child of the bookstore element.
//title[@lang]	Selects all the title elements that have an attribute named lang
//title[@lang='en']	Selects all the title elements that have a "lang" attribute with a value of "en"

Input1 (xml):

```
<html>
  <head>
    <meta charset = "utf-8">
    <title>Links</title>
  </head>
  <body>
    <h1>Here are my favorite sites</h1>
    <p><strong>Click on a name to go to that page</strong></p>
    <ul>
      <li><a href = "http://www.youtube.com">YouTube</a></li>
      <li><a href = "http://www.wikipedia.org">Wikipedia</a></li>
      <li><a href = "http://www.amazon.com">Amazon</a></li>
      <li><a href = "http://www.linkedin.com">LinkedIn</a></li>
    </ul>
  </body>
</html>
```

Input1 (pattern)

```
//h1/value
//body/li/value
//li[2]/a/value
a[@href]/attribute
p/value
```

Output1

Here are my favorite sites

Wikipedia

<http://www.youtube.com>

<http://www.wikipedia.org>

<http://www.amazon.com>

<http://www.linkedin.com>

Click on a name to go to that page

Input2 (xml)

```
<xml>
  <div class="r-ent">
    <div class="title">
      <a href="/bbs/Gossiping/P1.html">Title1</a>
    </div>
    <div class="date">11/10</div>
  </div>
  <div class="r-ent">
    <div class="title">
      <a href="/bbs/Gossiping/P2.html">Title2</a>
    </div>
    <div class="date">11/10</div>
  </div>
</xml>
```

Input2(pattern)

```
//div[@class='r-ent']/a/value
//div[@class='title']/a/value
//div[@class='date']/value
//div[@class]/attribute
```

Output2

```
Title1
Title2

Title1
Title2

11/10
11/10

r-ent
title
date
r-ent
title
date
```

注意事項:

1. xml 檔案內會在後方加入雜訊，因此需判斷原始 xml 已結束
2. /value 代表顯示 tag 的 value 值
3. /attribute 代表顯示 tag 某個 attribute 的值
4. 每筆測資中會有許多 pattern，每個 pattern 輸出之間需空行
5. 若 pattern 比對不到則輸出一空行
6. 原始 xml 符合 xml 格式
7. 需列出所有符合條件的結果

截止時間：

2021.12.15 23:59

繳交方式：

批改系統、Portal

作業系統：

Ubuntu 16.04

程式語言：

C or C++ (gcc version 9.4.0)

Command：

./hw2.exe case1.xml case1.pattern

Provide data：

mailto: tinin@saturn.yzu.edu.tw

title:[DSHW2] SID 測資提供

參考資料:

https://www.w3schools.com/xml/xpath_syntax.asp

<https://ithelp.ithome.com.tw/articles/10218625>

注意事項：

1. 傳值方式

```
int main(int argc, char* argv[])
{
    ifstream fin1, fin2;
    fin1.open(argv[1]);
    fin2.open(argv[2]);
}
```
2. 不要有 system("pause");
3. 遲交一天扣 10 分