Intro to XPath

Elli Mylonas
elli.mylonas@gmail.com

BærUt!
Sustainable Digital Scholarly Editions

This Presentation: http://tiny.cc/BaerUt-XPath

So you have an XML file!

You may want to:

- Process it
- Analyze it
- Improve the markup, update or refactor it
- Explore a file to understand how it has been encoded

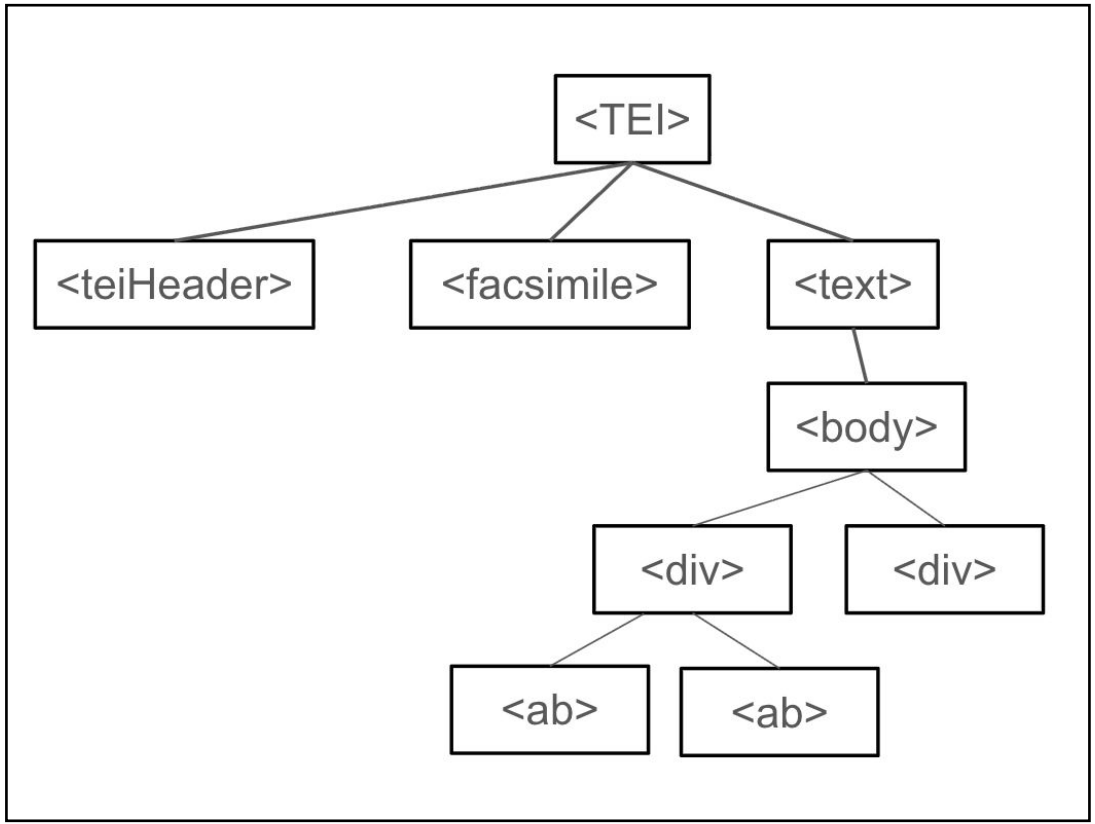
Before you can write a program to do any of this, you have to be able to locate and point to elements based on their position in the XML tree.

Examples

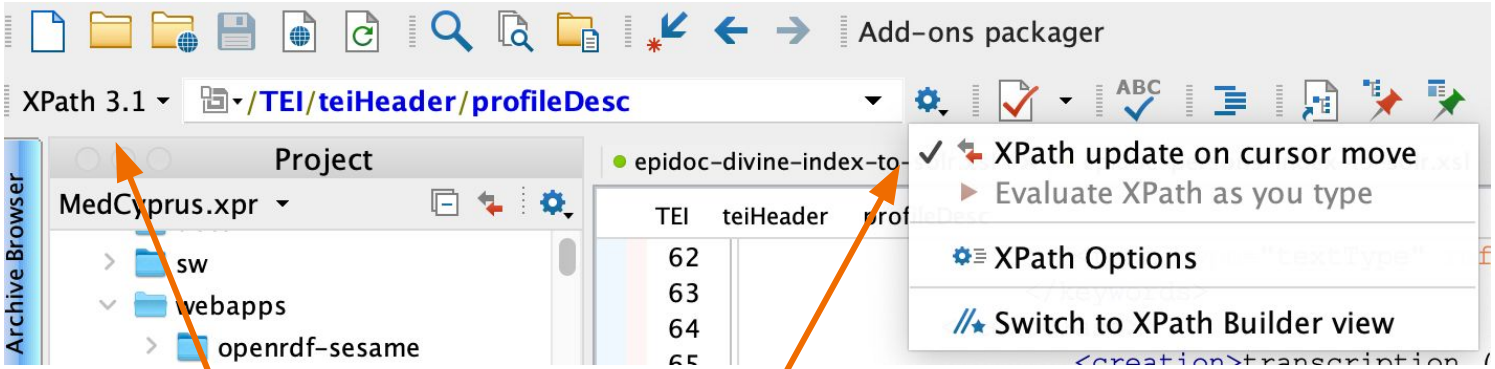
- Find all the chapter headings — to generate a TOC
- Find all the names of people and places — to generate an index
- Change all <p> to <ab> but only in
- Find all the chapters that have subsections
- Find all the personal names that have a @role attribute
- Find all <author> elements that are in the front matter
- Find the fourth line of each stanza

XPath is a query language designed for selecting nodes or node-sets from an XML document. It provides a way to navigate through the tree structure of an XML document and precisely identify and select specific elements, attributes, or other parts of the document.

```
<TEI>
    <teiHeader>
     ...
    </teiHeader>
    <text>
       <body>
         <div>
            <ab>...</ab>
            <ab>...</ab>
          </div>
          <div>
              ...
          </div>
        </body>
    </text>
</TEI>
```



Exploring XPath with Oxygen



The drop down menu on the left of the XPath box should be set to “XPath 2.0” or “XPath 3.1”.

The drop down menu next to the gear on the right of the XPath box should be set to “XPath update on cursor move.

Paths in XPath

Because elements contain other elements, the simplest way to locate an element is to walk down to it from the root of the XML tree.

```
/journal/entries/entry/date
```

```
//date
```

```
/journal/entries//date
```

```
//date/@when
```

Will all of these XPaths have the same result?

This will be familiar to you from your computer file system.

You can test these XPaths on the following file:

1. From Oxygen's **File** Menu, choose Open **URL**...
2. Paste in this URL:
   https://raw.githubusercontent.com/emylonas/LIS542A-Sources/refs/heads/main/Week5-XPath/AdamsXPath.xml
3. You can work on this file directly, or, use Save As... to save it to your computer.

XPath Predicates

It is useful to be able to constrain a path expression by adding information provides greater specification but isn't always part of the path.

```
//name[@role]
```

```
//name[@role="esquire"]
```

Compare with `//name/@role`

```
//p[pb]
```

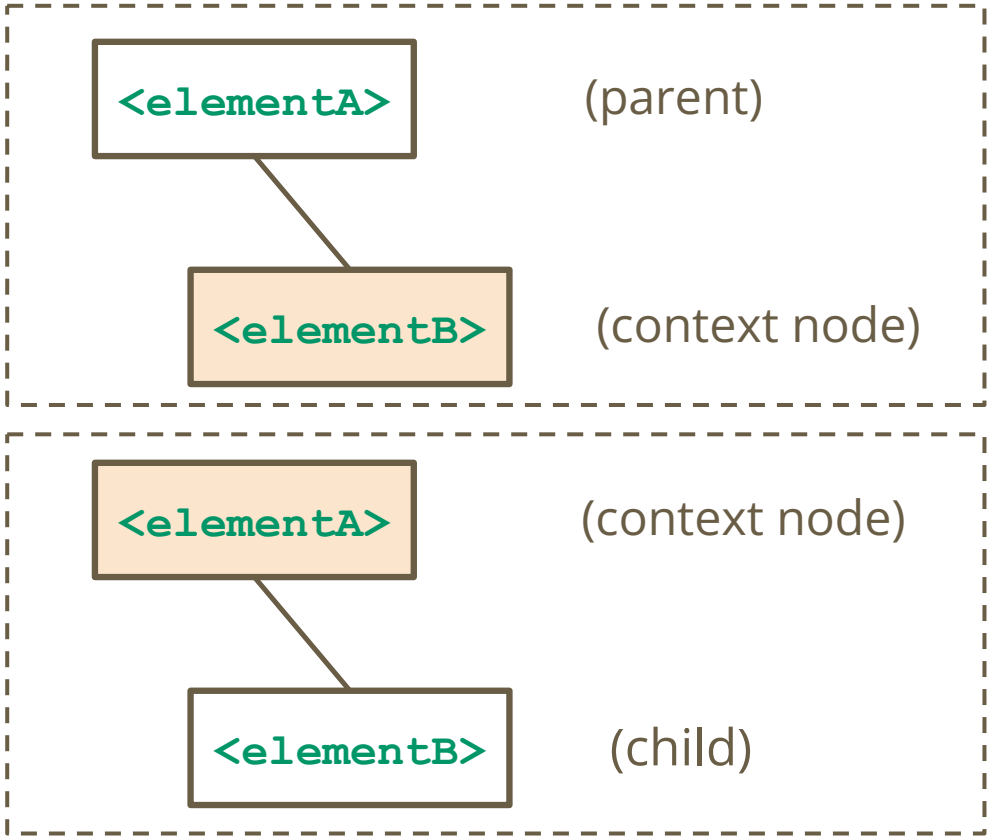
```
//p[2]
```

```
//p[2][pb]
```

XPath Axes

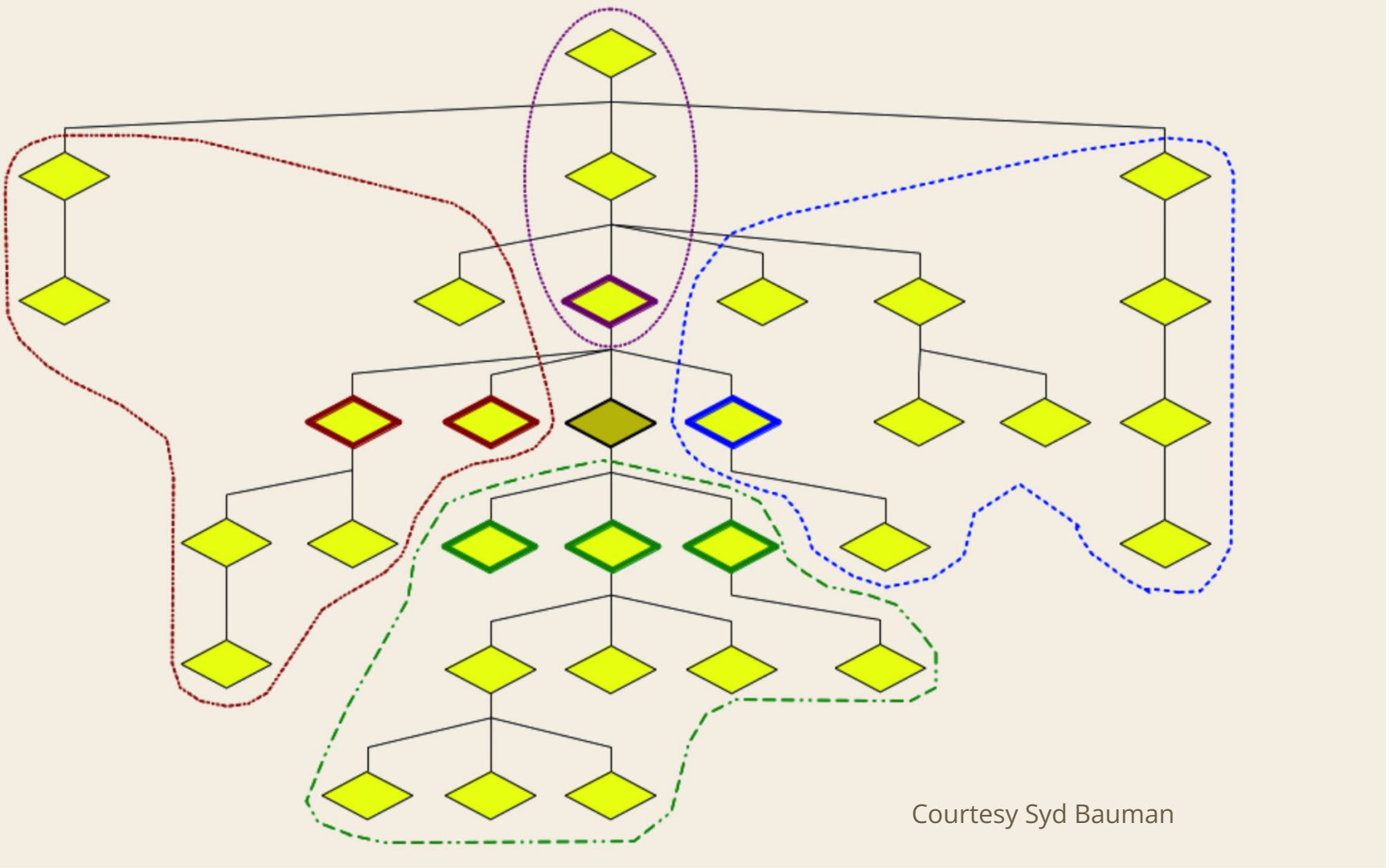
Nodes in the XML tree have relationships that can be used to specify navigation.

These are called *axes*.



Most Commonly Used Axes

child::	child of current node
descendant::	child or grandchild of current node
parent::	parent of current node
ancestor::	parent or ancestor of current node
preceding-sibling::	preceding current node, with same parent
preceding::	any node preceding the current node
following-sibling::	following current node, with same parent
following::	any node following the current node
self::	the current node



Courtesy Syd Bauman

Explore Axes

`//entry[child::date]` is the same as `//entry[date]`

`//name[parent::p[pb]]`

`//name[sibling::alternates]`

`//name[following-sibling::alternates]`

`//name[ancestor::metadata]` is the same as `//metadata//name`

XPath Functions

In addition to to specifying paths through the XML tree, XPath also has a series of functions that can further define sets of nodes.

not()

```
//name[not(@role)]
//p[not(pb)]
```

contains()

```
//name[contains(@role,'clergy')]
```

compare with

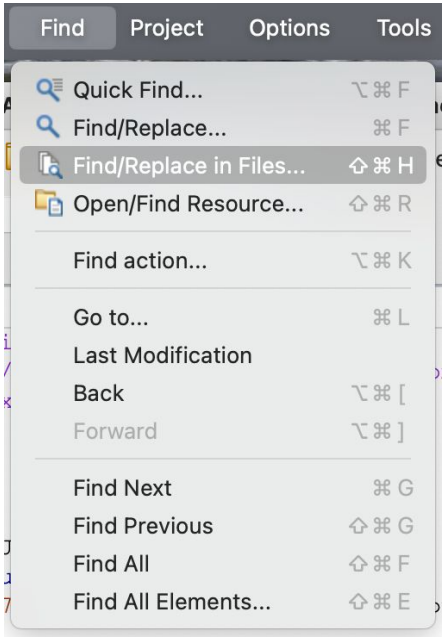
```
//@role[contains(.,'clergy')]
```

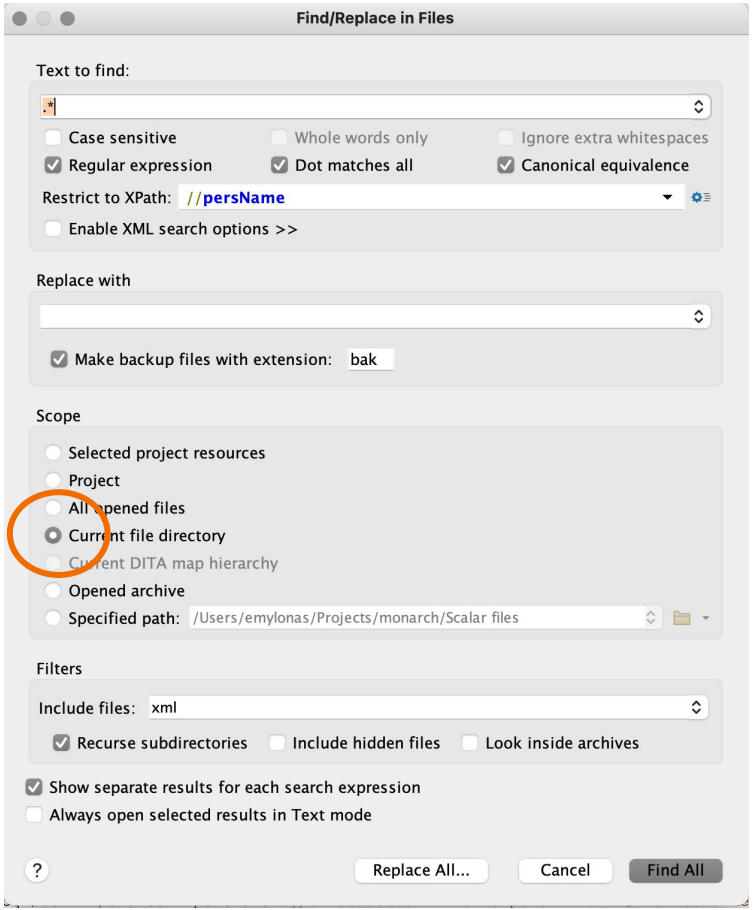
count()

```
//entry/count(p)
```

Search across all files in current directory

1. Select **`Find/Replace in Files`** from the **`Find`** menu.
2. Make sure to select **`Current File Directory`** in the Scope section of the dialog box.
3. If you are not looking for a particular string of text, put **`.*`** in the search box and select **`Regular Expression`** below it.





Further Reading

1. Birnbaum, David. “What Can XPath Do For Me”
   This is written in a slightly technical, formal way, and does a good job of laying out the basic concepts. http://dh.obdurodon.org/introduction-xpath.xhtm

2. Holmes, Martin. “XSLT” from the DariahTeach “TEI and XSLT” module.
   https://teach.dariah.eu/mod/page/view.php?id=454 View until 5:44 where the speaker starts to talk about XSLT.

3. W3Schools. “XPath Tutorial” Best used as a reference for syntax and functions.
   https://www.w3schools.com/xml/xpath_syntax.asp

Practice

Work through James Cummings' XPath introduction

https://tei-c.org/Vault/Talks/OUCS/2005-02/tei-oucs-IntroToXPath.pdf