

# *XPath node predicates*

Martin Holmes

## XPath Node Predicates

- You use paths and axes in XPath to arrive at specific nodes in your XML.
- You use predicates to further filter or test those nodes.
- Only nodes which satisfy the predicate will be selected.
- Predicates **follow** the step they apply to, and use **square brackets**.
- `/TEI/text/body/div` means "all the `<div>` nodes which are children of `<body>`".
- `/TEI/text/body/div[@type='chapter']` means "only those `<div>` nodes which are children of `<body>` **and also have a type attribute with the value "chapter"**".

## XPath Predicates: some examples

- `//div[head]`
- = all `<div>` nodes which have a direct child `<head>` element.
- `//div[contains(head, "Scene")]`
- = all `<div>` nodes which have a direct child `<head>` element which contains the text "Scene".
- `//div[1]`
- = all `<div>` nodes which are the first `<div>` node in their parent element.
- `//div[position() = last()]`
- = all `<div>` nodes which are the last `<div>` node in their parent element.

## XPath Predicates: Tasks (1)

Using the **places.xml** file here: <http://web.uvic.ca/~mholmes/dhoxss2013/examples/places.xml> and the XPath console in oXygen, find the following information:

- How many place entries are mapped as polygons? (Hint: a polygon location has more than 2 <geo> elements.)
- How many are single points?
- How many place entries contain external links (<ref type="external">)?

## XPath Predicates: Tasks (1) answers

Using the **places.xml** file, and the XPath console in oXygen, find the following information:

- **Q:** How many place entries are mapped as polygons? (Hint: a polygon location has more than 2 <geo> elements.)
- **A: 77**

```
count(//place[count(location/geo) gt 2])
```

- **Q:** How many are single points?
- **A: 110**

```
count(//place[count(location/geo) eq 1])
```

- **Q:** How many place entries contain external links (<ref type="external">)?
- **A: 105**

```
count(//place[descendant::ref[@type='external']] )
```

(Note the nested predicates!)

## XPath Predicates can be chained

The following examples work on the Hamlet XML file here: <http://web.uvic.ca/~mholmes/dhoxss2013/examples/hamlet.xml>

- ```
//div[parent::div][count(descendant::l) gt 200]
```
- = all scenes containing more than 200 lines.
- ```
//div[parent::div][descendant::sp/@who='#Horatio' ]
```
- = all scenes in which Horatio speaks.

## XPath node predicates

---

- `//sp[position()=last()][@who='#Hamlet']`
- = speeches by Hamlet which are the last speeches in their scene.
- Supplementary question: how would you find the last speech in the whole play? Hint: you could use the `following::` axis and the `not()` function, or you could use a long path with several instances of `position()=last()`.

The Hamlet file actually has lots of `<ab>` elements in speeches as well as `<l>` elements, so this line count stuff is not a true reflection of the length of speeches. It might be worth telling the students there's something wrong with it, and asking them to look at the XML source to see if they can figure out what it is. Then you could amend the first item in this slide, and some of the ones in the next slide, to account for this.

## XPath Predicates can be nested

- `//div[child::div][descendant::sp[@who='#Ophelia']]`
- = all acts in which Ophelia speaks.
- `//div[parent::div][count(descendant::sp[@who = '#Hamlet']) > 30]`
- = all scenes in which Hamlet has more than 30 speeches.
- `//sp[descendant::l[descendant::stage]]`
- = all speeches which contain lines which contain stage directions

## XPath Predicates: Tasks (2)

Using the **places.xml** file, and the XPath console in <oXygen/>, find the following information:

- How many place entries mention the name whose `@key` value is "kellett"?
- How many place entries have a bibliography with more than two items?
- How many place entries which are single points contain external links?

## XPath Predicates: Tasks (2) answer 1

- How many place entries mention the name whose @key value is "kellett"?
- `count(//place[descendant::name[@key='kellett']])`

(The answer is 9.)

## XPath Predicates: Tasks (2) answer 2

- How many place entries have a bibliography with more than two items?
  - `count(//place[count(descendant::bibl[parent::listBibl]) > 2])`
- (The answer is 12.)
- Note: if we omit the [parent::listBibl] predicate, we will be counting bibl elements which appear outside the bibliography list, in the body of the explanatory text.

## XPath Predicates: Tasks (2) answer 3

- How many place entries which are single points contain external links?
  - `count(//place[descendant::ref[@type='external']][count(./location/geo) = 1])`
- (The answer is 58.)

## XPath predicates: summary

- Predicates **filter** the results of an XPath.
- They **follow** the step they apply to.
- They're enclosed in **square brackets**.
- They can be **chained**.
- They can be **nested**.