

XML and queries on XML xpath, xquery

Grace Swenson Hollis



XML stands for

Extensible Markup Language



XML was first released...

Started in 1996 but first published February 10th, 1998.

Bonus: who invented it?

World Wide Web Consortium



So what is it?

- Derived from SGML (Standard Generalized Markup Language), but simpler to use
- Documents have tags giving extra information about sections of the document
- Extensible (old versions still work!!!)
- Users can add new tags
- Designed to store and transport data



How is it different than HTML?

XML and HTML were designed with different goals:

- XML was designed to carry data - with focus on what data is
- HTML was designed to display data - with focus on how data looks
- XML tags are not predefined like HTML tags are
- XML is often used to store/transfer data, while HTML is used to present it

[Thanks w3schools](#)



XML Structure

- Tag: label for a section of data
- Element: section of data beginning with and ending with matching
- Elements must be properly nested

Proper nesting ... `<word>...<character> A</character></character...<word>`

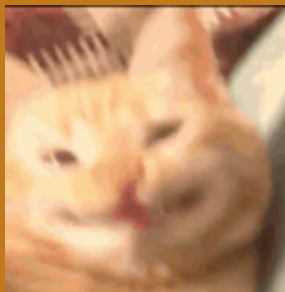
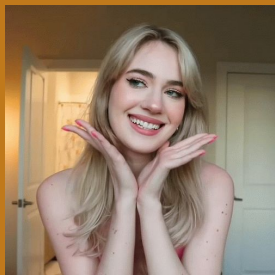
Improper nesting ... `<word>...<character> A</character></word></character>`

- Formally: every start tag must have a unique matching end tag, that is in the context of the same parent element.
- Every document must have a single top-level element



Example of Nesting

```
<Bryn_Mawr_college>  
  <people>  
    <faculty>  
      <Professor> Geoffrey Towell</Professor>  
    </faculty>  
  </people>  
  <building>  
    <building_name> Park</building_name>  
  </building>  
</Bryn_Mawr_college>
```



Geoff quote break

Cool! Excellent!
And other words like that





Escape Sequences

- `<`; represents "<";
- `>`; represents ">";
- `&`; represents "&";
- `'`; represents "'";
- `"`; represents '"'



Attributes

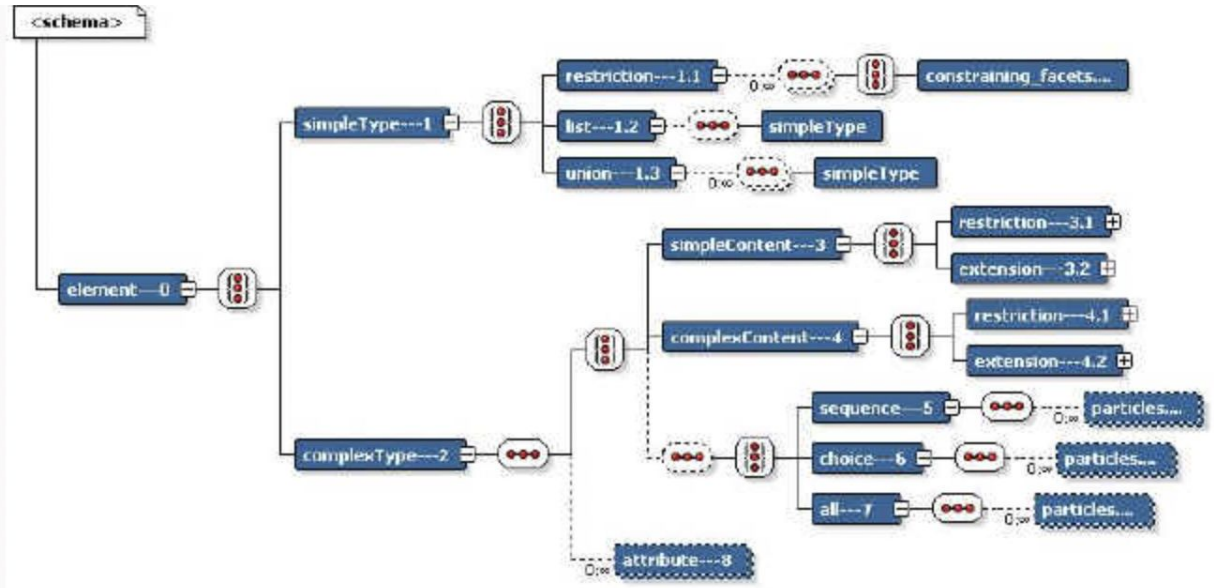
- Must be quoted, double or single
 - Ex: `<person grade= "junior">`
- Can't be expanded on in the future
- Can't contain tree structure
- Can't contain multiple values



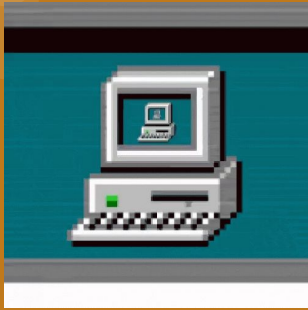
Document Schema

Document Type Definition (DTD)

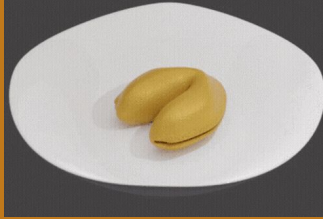
XML Schema



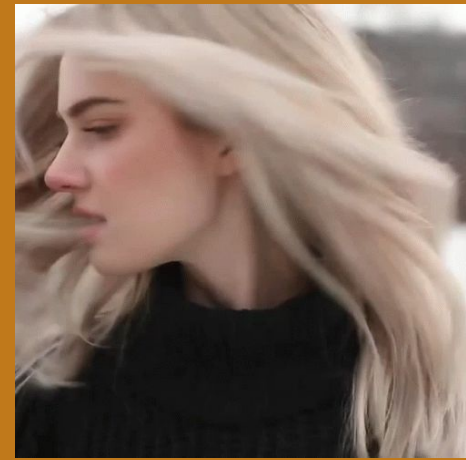
[Link](#)



Boolean Algebra



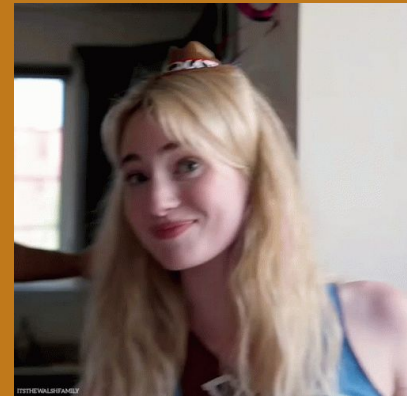
Geoff quote break



When you can get in a
computer science joke,
you really should



Me when I ask someone on
StackOverflow what an error
message means and they reply
"it means what it says"



Xpath and Xquery





Xpath

- XPath is used to address (select) parts of documents using **path expressions**
- A path expression is a sequence of steps separated by "/"
- Result of path expression: set of values that along with their containing elements/attributes match the specified path
- The initial "/" denotes root of the document (above the top-level tag)
- Path expressions are evaluated left to right
 - Each step operates on the set of instances produced by the previous step
- Selection predicates may follow any step in a path, in []
- Attributes are accessed using "@"
 - E.g. /university-3/course[credits >= 4]/@course_id



Operators

Operator	Description	Example
	Computes two node-sets	//book //cd
+	Addition	6 + 4
-	Subtraction	6 - 4
*	Multiplication	6 * 4
div	Division	8 div 4
=	Equal	price=9.80
!=	Not equal	price!=9.80
<	Less than	price<9.80
<=	Less than or equal to	price<=9.80
>	Greater than	price>9.80
>=	Greater than or equal to	price>=9.80
or	or	price=9.80 or price=9.70
and	and	price>9.00 and price<9.90
mod	Modulus (division remainder)	5 mod 2

Thanks
[W3schools](https://www.w3schools.com/)



Xpath Functions and Features

- `count()`: count the number of elements in the set generated by a path
- Boolean `and()` and `or()`
- `false()`, `true()`, `boolean()`
- `id()`
- `last()`
- `name()`
- `sum()`
- `//` skip multiple levels of nodes
- `doc(name)` returns the root of a named document [Here is more](#)



Geoff quote break

Why would I want to do rainbow when I could do dry erase markers and ruin two things?





Xquery

- XQuery is derived from the Quilt query language, which itself borrows from SQL
- uses a **for ... let ... where ... order by ...result ...** syntax

Xquery

SQL

for

From

where

Where

order by

Order by

result

Select

Let

No equipment



Joins

- Similar to SQL joins, kind of

```
for $c in /university/course,
```

```
    $i in /university/instructor,
```

```
    $t in /university/teaches
```

```
where $c/course_id= $t/course id and $t/IID = $i/IID
```

```
return { $c $i }
```

- You could also use xpath

```
for $c in /university/course,
```

```
    $i in /university/instructor,
```

```
    $t in /university/teaches[ $c/course_id=
```

```
        $t/course_id
```

```
and $t/IID = $i/IID]
```

```
return { $c $i }
```



sources

MozDevNet. (n.d.). *Functions - xpath: MDN*. MDN Web Docs. <https://developer.mozilla.org/en-US/docs/Web/XPath/Functions>

oracle. (n.d.). XML schema: Understanding structures. <https://www.oracle.com/technical-resources/articles/srivastava-structures.html>

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). *Database system concepts*. McGraw-Hill Education.

<https://db-book.com/slides-dir/PDF-dir/ch30.pdf>

W3Schools. (n.d.). XML tutorial. <https://www.w3schools.com/xml/>

What is XML? - XML file explained - AWS. Amazon Web Service. (n.d.). <https://aws.amazon.com/what-is/xml/>

Wikimedia Foundation. (2024, April 13). *XML*. Wikipedia. <https://en.wikipedia.org/wiki/XML>



**BUT WAIT
THERE'S MORE**

