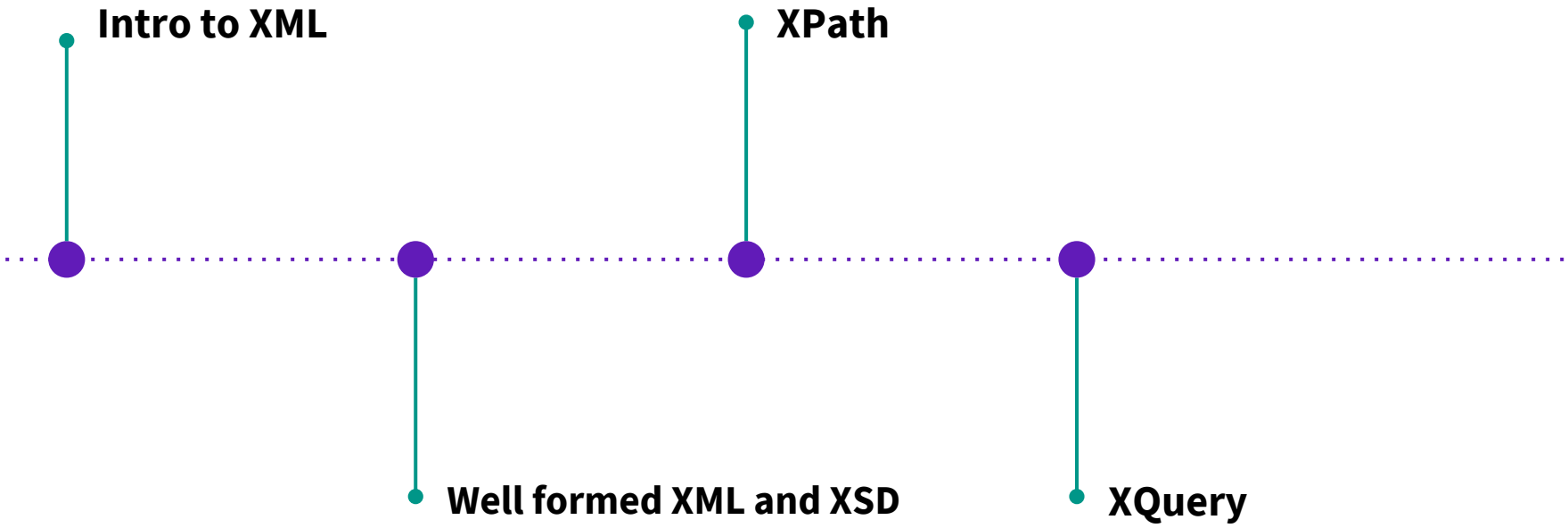
XML, XQuery and Xpath

Massimiliano Luca
massimiliano.luca@unitn.it

Nicolò Alessandro Girardini
nicolo.girardini@studenti.unitn.it

Introduction to Service Design and Engineering
Tech Lab: 7th November 2018

Agenda



Tools

XML & XSD

- Text editor (Atom or Sublime are good)
- Free Formatter - XML XSD Validator (freeformatter.com/xml-validator-xsd.html)

XPath

- Free Formatter - XPath Tester (freeformatter.com/xpath-tester.html)

XQuery

- XQuery Tester (www.xpathtester.com/xquery)

Introduction to XML and XSDs

eXtensible Markup Language

- Standard for data representation and data exchanging
- Document format is similar to HTML
  - Tags describe contents instead of formatting
  - UTF-8 compliancy

XML Basics

```xml
<Universities>
  <University name="University of Trento" longitude="42.64" latitude="68.93">
    <Courses>
      <Course name="Knowledge and Data Integration" code="145324">
        <TeachingTeam>
          <Professor>
            <FirstName>Fausto</FirstName>
            <LastName>Giunchiglia</LastName>
          </Professor>
          <Assistant>
            <FirstName>Mattia</FirstName>
            <LastName>Fumagalli</LastName>
            <Title>Ph.D.</Title>
          </Assistant>
        </TeachingTeam>
      </Course>
    </Courses>
  </University>
</Universities>
```

Basic constructs

- Tagged elements
- Attributes
- Text

Structural requirements

- Single root
- DTD/XSD compliancy (if DTD/XSD is defined)
- Unique attributes within elements

XML Visualization

- To properly show XML as HTML use:
  - CSS (Cascading Stylesheets)
  - XSL (eXtensible Stylesheet Language)

Relational Schemas vs. XML

	Relational schema	XML
Structure	Table	Hierarchical Tree
Schema	Fixed in advance (ER / DB structure)	Flexible
Query	Simple human readable languages (SQL)	There are languages but are not that readable and simple
Ordering	None	Implied
Implementation	Native	Add-On

XML or Relational, which to choose?

- The Professor want to store some simple information to track who took this class (student id and name) and the mark of each student
- You are a web developer and you are developing a website for a big company. You want to store news, events, open positions, ...

XML or Relational, which to choose?

- The Professor want to store some simple information to track who took this class (student id and name) and the mark of each student

  A relation db: in this case the format of the data does not change so it is preferable to have a schema which is well defined

- You are a web developer and you are developing a website for a big company. You want to store news, events, open positions, ...

  XML: it is a more dynamic environment and a relational schema is to fixed to model this scenario

XSDs

- XML Schema Descriptor
- XML file with a specific grammar to define the elements (new tags), attributes, nesting, ordering, number of occurrences
- There are also IDs and REFs, which are nothing more than pointers

XSD Basic Syntax

Basic components:

- <xs:schema> is the root node, used to define the namespace (elements/types)
- <xs:element> to define elements
- <xs:attribute> to define attributes
- <xs:sequence> to define an order

To specify the characteristics we add attributes in nodes, like *name* and *type*

NOTE: Better explanation in the cheatsheet

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="note">
    <xs:complexType>
      <xs:element name="to" type="xs:string"/>
      <xs:element name="from" type="xs:string"/>
      <xs:element name="heading" type="xs:string"/>
      <xs:element name="body" type="xs:string">
        <xs:attribute name="lang" type="xs:string" default="EN"/>
      </xs:element>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Why Use a Schema?

With Schema	Without Schema
● Programmers that want to use the resource can assume a structure ● If there is a structure, XSL or CSS can be used to enhance the visualization ● There is some kind of specification and so the exchange of data can be simpler to implement	● Flexibility ● XSDs may contain errors and as far as XSDs are used to valid XML files this is a huge problem (a.k.a. XSDs can be messy) ● Same for DTDs

Practice

Exercise 1 - XML Construction

Build an XML file with those characteristics:

- `students` is the root node
- There will be two `student` subnodes
- Each student node will have as attributes: `id`, `name`, `age`, `major` (course path)
- Each student will have a `results` subnode: `result` will be the name of the subelement, with attributes `course` and `grade` (optional)
- The first student is: 198449, Massimiliano, 23, Computer Science. He has taken the Math course with grade C-
- The second student is: 203265, Nicolò, 23, Computer Science. He took the Math course with grade A and the Data Analysis course, but we don’t know the grade

Exercise 1 - XML Construction

```xml
<students>
  <student id="198449" name="Massimiliano" age="22" major="Computer Science">
    <results>
      <result course="Math" grade="C-"/>
    </results>
  </student>
  <student id="203265" name="Nicolò" age="23" major="Computer Science">
    <results>
      <result course="Math" grade="A"/>
      <result course="Data Analysis"/>
    </results>
  </student>
</students>
```

Exercise 2 - Structured XML

Working on the previous XML file:

- `students` has to become a more structured node
- Leave `id` as attribute and `results` as it is
- Each student node will have the other attributes as subnodes

Exercise 2 - Structured XML

```xml
<students>
  <student id="198449">
    <name>Massimiliano</name>
    <age>22</age>
    <major>Computer Science</major>
    <results>
      <result course="Math" grade="C-"/>
    </results>
  </student>
  <student id="192414">
    <name>Nicolo'</name>
    <major>Computer Science</major>
    <major>Secutiy and Network</major>
    <results>
      <result course="Math" grade="A"/>
      <result course="Data Analysis"/>
    </results>
  </student>
</students>
```

Exercise 3 - XSD Definition

Provide a possible XSD

given the following XML

```xml
<students>
  <student name="David"></student>
  <student name="Craig"></student>
  <student name="Erik"></student>
</students>
```

Exercise 3 - XSD Definition

Provide a possible XSD

given the following XML

```xml
<students>
  <student name="David"></student>
  <student name="Craig"></student>
  <student name="Erik"></student>
</students>
```

```xml
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="students">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="student" maxOccurs="unbounded" minOccurs="0">
          <xs:complexType>
            <xs:attribute type="xs:string" name="name" use="optional"/>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Exercise 4 - XSD Match

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="person">
    <xs:complexType>
      <xs:attribute type="xs:string" name="name" use="optional"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="food">
    <xs:complexType>
          <xs:attribute type="xs:string" name="name" use="optional"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="eats">
    <xs:complexType>
          <xs:attribute type="xs:string" name="diner" use="optional"/>
          <xs:attribute type="xs:string" name="dish" use="optional"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="meal">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="person" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element ref="food" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element ref="eats" maxOccurs="unbounded" minOccurs="0"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Is this XML matched?

```
<meal>
  <person name="Alice"/>
  <food name="salad"/>
  <eats diner="Alice" dish="salad"/>
  <person name="Bob"/>
  <food name="salad"/>
  <eats diner="Bob" dish="salad"/>
  <person name="Carol"/>
  <food name="sandwich"/>
  <eats diner="Carol" dish="sandwich"/>
</meal>
```

```
<meal>
  <person name="Alice"/>
  <person name="Bob"/>
  <person name="Carol"/>
  <person name="Dave"/>
  <food name="salad"/>
  <food name="turkey"/>
  <food name="sandwich"/>
  <eats diner="Alice" dish="turkey"/>
  <eats diner="Bob" dish="salad"/>
  <eats diner="turkey" dish="Dave"/>
</meal>
```

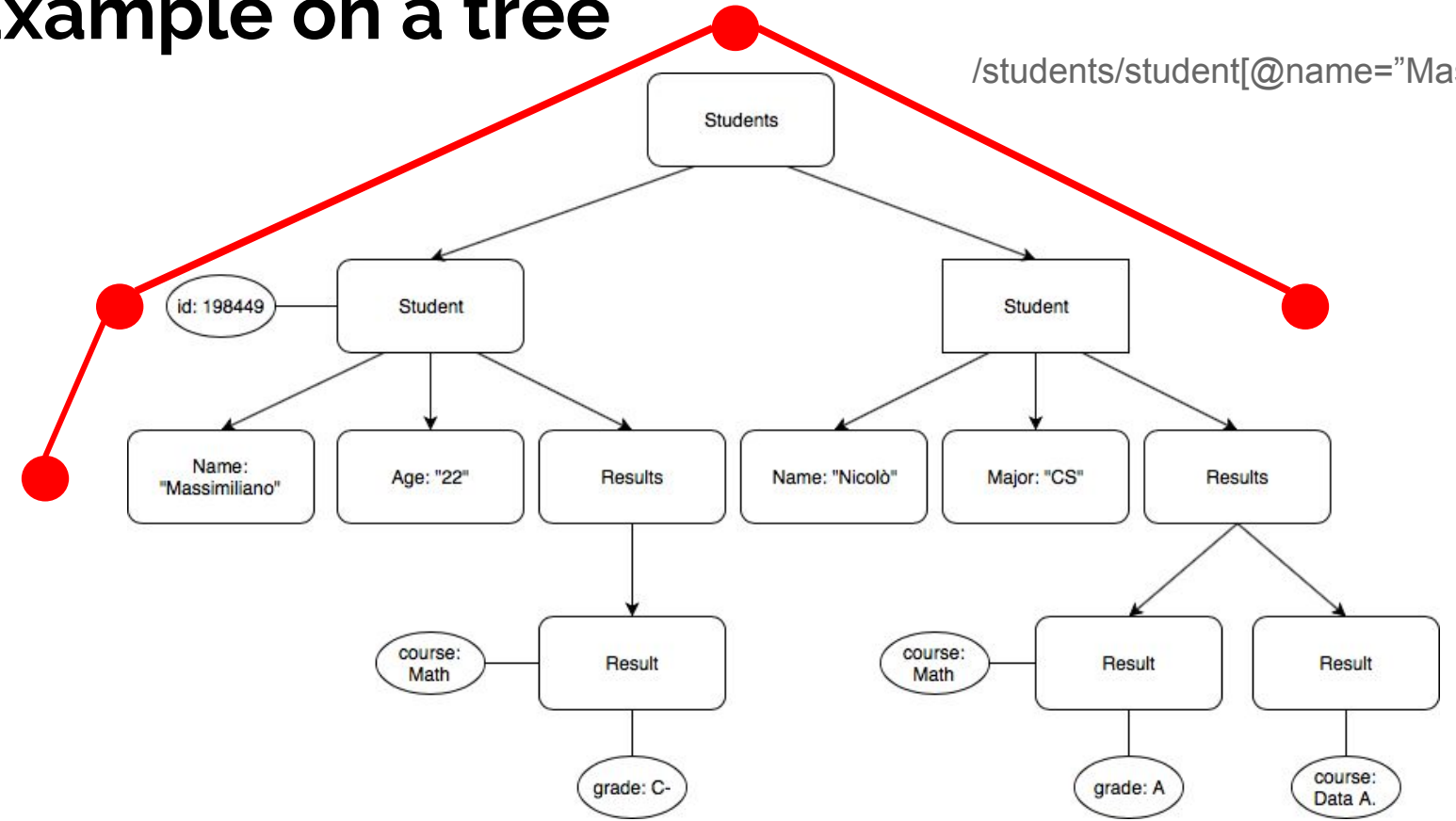
XPath

XPath

- Tool to navigate through XML documents and select objects
- It works like path navigation in OSs
- Not mature as querying relational databases
- No underline algebra (implies also less intuitive query)
- XPath allows to look for specific paths in the underlying tree and to set up some simple conditions

Example on a tree

/students/student[@name=”Massimiliano”]



XPath Basic Operators

Basic constraints:

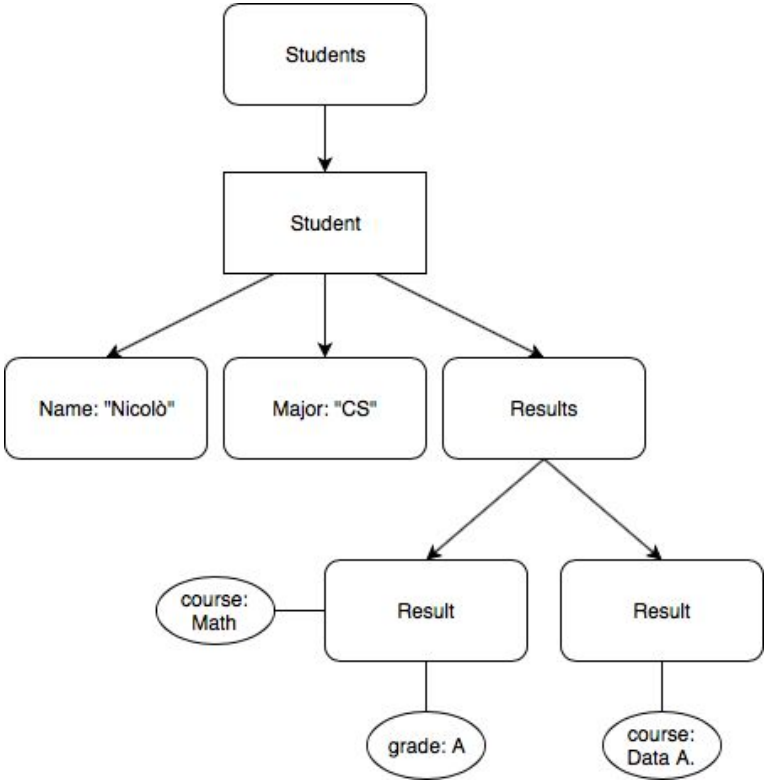
/ - to navigate the root element (it is also the separator)

<element_name> - to navigate through a specific element (as in SQL, * can be used)

@<attribute_name> - to access the value of an attribute

// - to match any descendant of the current node

[<cond>] - to evaluate a condition at a certain level



XPath Examples

Access to “results”

//results

/students/student/results

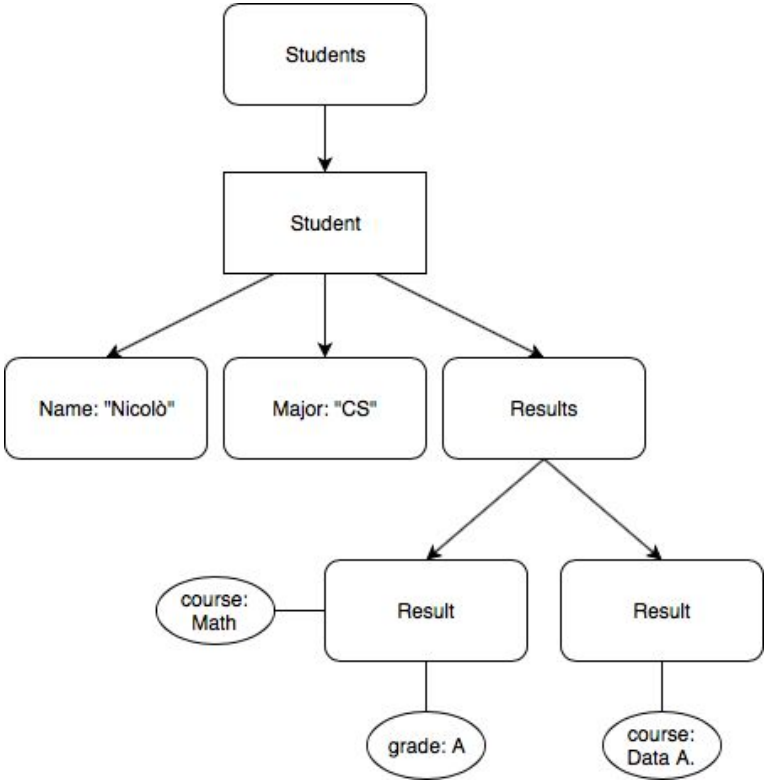
Access to a specific student

/students/student[1]

/students/student[last()]

Conditions

/students/student[name="@Nicolò"]/results



XPath Examples

Access to “results”

//results

/students/student/results

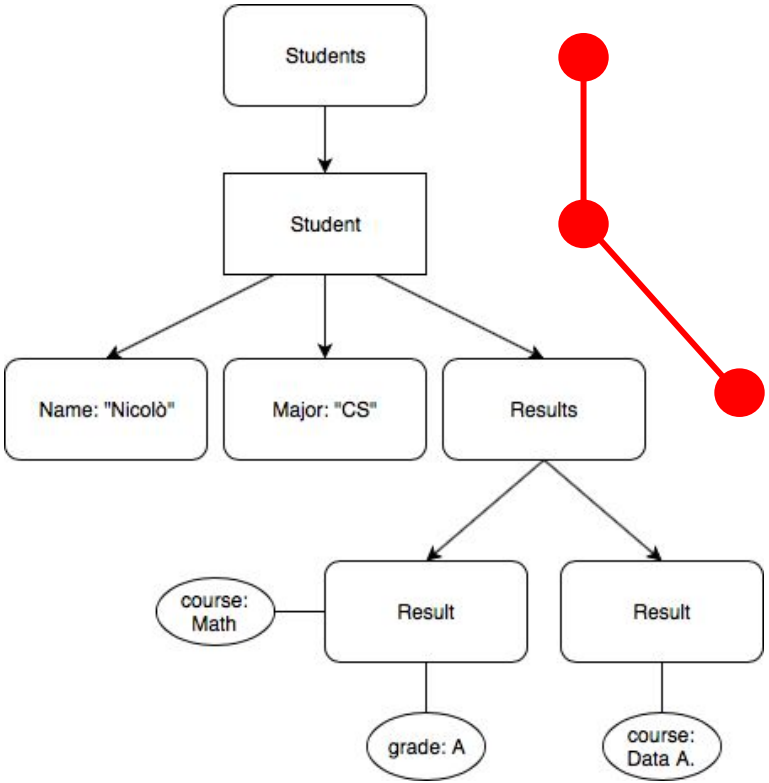
Access to a specific student

/students/student[1]

/students/student[last()]

Conditions

/students/student[name="@Nicolò"]/results



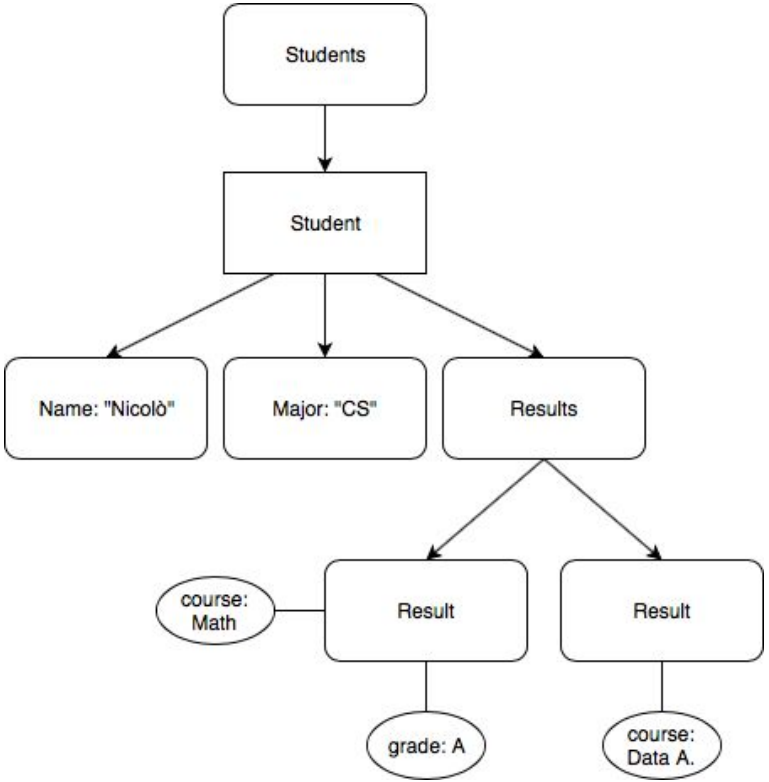
XPath Examples

Access to all the results with the attribute grade

//result[@grade]

Filter on the attribute

//result[@grade="A"]

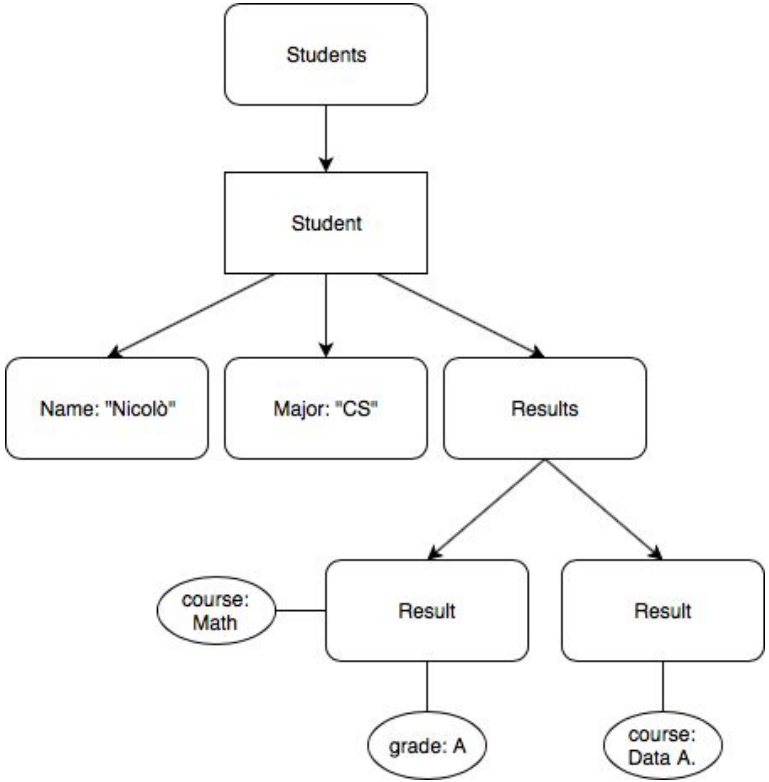


XPath Axes and Built-In Functions

Advanced constraints:

- built-in functions
  - contains(string1, string2)
  - name()
- Axes
  - parent::
  - descendants::
  - self::
  - ancestor::
  - … other 11

NOTE: More details in the cheatsheet



Practice

“Presentations aren't efficient without cats”
- Lex Fridman, research scientist @ MIT

Exercise - XPath Queries

1. Return all the student older than 20
2. Return all the students that are studying computer science
3. Return all the student that took the “Data Analysis” class

```
<students>
  <student id="198449">
    <name>Massimiliano</name>
    <age>22</age>
    <major>Computer Science</major>
    <results>
      <result course="Math" grade="C-"/>
    </results>
  </student>
  <student id="192414">
    <name>Nicolo'</name>
    <major>Computer Science</major>
    <major>Secutiy and Network</major>
    <results>
      <result course="Math" grade="A"/>
      <result course="Data Analysis"/>
    </results>
  </student>
</students>
```

Exercise - Guided Solutions

1. //student[age>20] **or** /students/student[age>20]

2. //student[major="Computer Science"]/parent::*

Exercise - All Solutions

1. //student[age>20] **or** /students/student[age>20]
2. //student[major="Computer Science"]/parent::*
3. //student/results/result[@course="Data Analysis"]/ancestor::student

XQuery

XQuery

- Expression language (compositional)
- Work on sequences of elements
- Return sequences of elements
- FLOWR expression
    - For
    - Let
    - Order
    - Where
    - Return

XPath VS XQuery

Access to “results”

//results

/students/student/results

Access to a specific student

//student/age>20/parent::student/major

Can’t be ordered

Access to “results”

/students/student/results

Access to a specific student

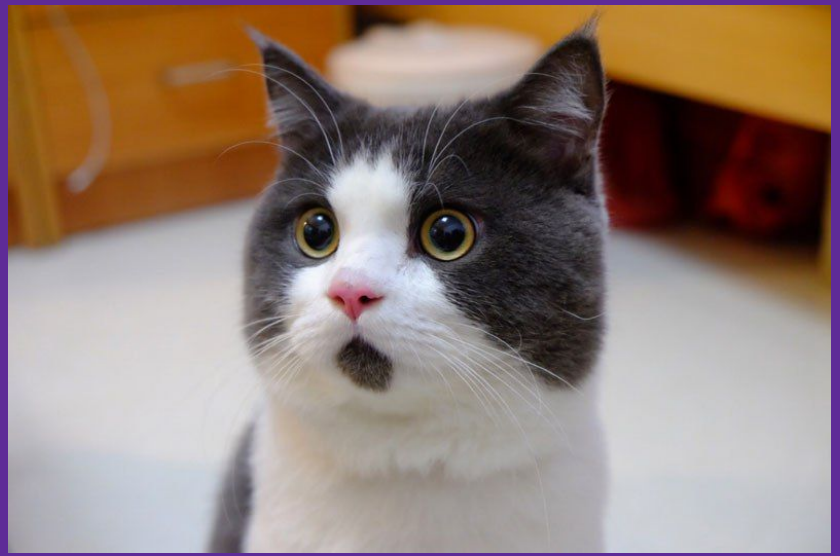
for $x in /students/student

where $x/age>20

order by $x/age

return $x/major

The Truth



Real World Operations

- XMLs won’t be read often by humans: it is usually used to transfer data and to interface applications/services
- Generation of schemas is mostly done with tools
- Every major language/framework has libraries to manage XML files and to navigate through them with tools such as XPath and XQuery

- Java: `javax.xml.parsers; javax.xml.xpath; javax.xml.xquery`
- Python: `libxml2`
- C++: `libxml++`

Homeworks

Navigate to github.com/MassimilianoLuca/Introduction-to-XML/ and download “countries.xml” :

1. Generate the XSD file for the document given (There are tools to do it)
2. Return the names of all countries where a city in that country contains more than one-third of the country's population (With XPath)
3. Return the average number of languages spoken in countries where Russian is spoken (With Xpath)
4. Return the names of all countries whose name textually contains a language spoken in that country. For instance, Uzbek is spoken in Uzbekistan, so return Uzbekistan (With XQuery)

Expected solution values can be found on github - expected_solution.md

Contact

Massimiliano Luca
massimiliano.luca@unitn.it
massimilianoluca.github.io

Nicolò A. Girardini
nicolo.girardini@studenti.unitn.it

Hope you enjoyed

Please, fill up this form

goo.gl/SkdN8V