



Introduction to Database Systems: CS312

XML for Unstructured Data

Oliver Bonham-Carter

11 March 2019

What is XML?

Extendable Markup Language

Introduction

Code

XML Vs. HTML

Elements

Family of Nodes

Common Mistakes

Structure of Language

Let's Code

The Future!



- Unstructured versus structured database
 - Unlike SQL Data held in one major table
 - XML data may be spread-out across a large document
- XML language is flexible
 - Applied to diverse data types, constructions and forms

What Does XML Code Look Like?

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!

Partial XML Example

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Class</to>
  <from>Community</from>
  <heading>Reminder</heading>
  <body>Look after your health, everyday!</body>
</note>
```

Displaying...

Note

To: Class

From: Community

Look after your health, everyday!

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!



- XML
 - Data-centric
 - Designed to *store*, *contain* and *exchange* data
 - Contains many features for working with data
 - Features for making data useful.
 - Contains integrity constraints – enables the data to be checked for types of “correctness” (i.e., rules to determine how types of data are entered, accepts numbers or text only, etc.)

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!



- HTML
 - Display-centric
 - Designed to *display* data
 - For manipulating text attributes
 - Contains few (or *no*, depending on who you ask) tools for manipulating data.
 - Features for making data look nice.

- All XML documents have headers to declare XML syntax.
- First Line: `<?xml version="1.0" encoding="UTF-8"?>`
- Documents have one (and only one) root element
- Environments for data entries `<tag>` followed by `</tag>`
- Closers: Note the closing element, `</start>`

Partial XML Example

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Class</to>
  <from>Community</from>
  <heading>Reminder</heading>
  <body>Look after your health, everyday!</body>
</note>
```

XML: A Family Affair

Introduction

Code

XML Vs.
HTML

Elements

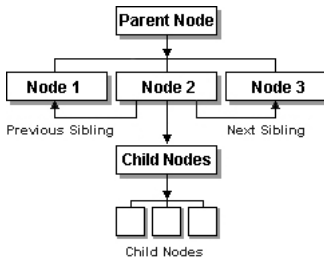
Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!



- XML documents must have only one root
- The elements under a root are its *children* .
- The elements may be considered to play the same roles as the *attributes* from SQL since they hold data values.
- All the elements under the root are siblings.
- Note that there are no double-quotes when defining the values of children.

Flexibility

We design own elements (using rules) to define environments

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!

- We wish to describe our Pets and plants
 - A Dog called, *Flint*
 - A Cat called, *Tiger*
 - A Fish called, *Wanda*
 - A Plant called, *Plantacious*
- How to put this data in a XML?

```
<?xml version="1.0" encoding="UTF-8"?>
<PetsAndPlants> <!--this is a root element-->
    <Dog> Flint </Dog> #this is a child element
    <Cat> Tiger </Cat>
    <Fish> Wanda </Fish>
    <Plant> Plantacious </Plant>
</PetsAndPlants>
```


Common Mistakes: Case Sensitive For Elements

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!

```
<?xml version="1.0" encoding="UTF-8"?>
<PetsAndPlants> <!--this is a root element-->
    <dog> Flint </Dog> <!--this is a child element-->
    <Cat> Tiger </Cat>
    <Fish> Wanda </Fish>
    <Plant> Plantacious </Plant>
</PetsAndPlants>
```

- Open this XML file in Firefox to see where the error is found.
- Mismatching element *pairs*

XML Parsing Error: mismatched tag. Expected: </dog>.
Location: file:///Users/obonhamcarter/Desktop/rootAndChild.xml
Line Number 3, Column 21:

```
<dog> Flint </Dog> <!--this is a child element-->
-----^
```

Common Mistakes: No Root Element!

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- <PetsAndPlants>    commented out r-->
    <Dog> Flint </Dog> <!--this is a child element-->
    <Cat> Tiger </Cat>
    <Fish> Wanda </Fish>
    <Plant> Plantacious </Plant>
<!-- </PetsAndPlants> -->
```

- Open this XML file in Firefox to see where the error is found.
- Parsing problems: No *root* to connect the data

XML Parsing Error: junk after document element
Location: file:///Users/obonhamcarter/Desktop/noRoot.xml
Line Number 4, Column 4:

```
    <Cat> Tiger </Cat>
    ---^
```

Designing Elements

Introduction

Code

XML Vs.
HTML

Elements

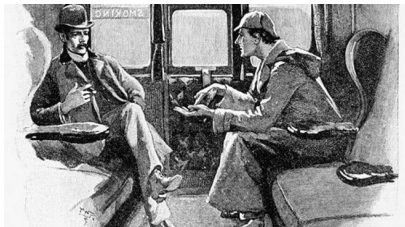
Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!



“Elementary, My Dear Watson.”
(BTW: Sherlock Never Said This)

- HTML has arbitrarily designed *elements* – we cannot make them up for our own data types.
- In XML, elements are designed according to a few rules
 - Names may contain numbers, letters, and letter-characters
 - Cannot start with a number or punctuation characters
 - Cannot start with letters, “XML” of any case.
 - Cannot contain spaces

Getting to the Data: A Python Application

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!

- How to get to particular information in XML files?
 - Actually, Google calls this format, *KLM*
 - https://developers.google.com/kml/documentation/kml_tut
- Parsing with Python is a good way...

Part of File: cityData.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document>
  <Placemark>
    <name>Denver, CO</name>
    <description>Located in Colorado</description>
    <Point>
      <coordinates>3944N,10459W</coordinates>
      <!-- lat, long -->
    </Point>
  </Placemark>
```

Consider this...

Please see the *sandbox* file for `cityData.xml`

Introduction

Code

XML Vs.
HTML

Elements

Family of
Nodes

Common
Mistakes

Structure of
Language

Let's Code

The Future!



THINK

- Can you create a similar program to parse other XML files?
- When would this tool be most useful to you?