# Analyzing Structured Data in Python

## Week 2: ElementTree

18 October 2020

CDCS Python Course Series

Instructor: Lucy Havens

# Course Topics

Week 1: Pandas for CSV data

**Week 2: ElementTree for XML data**

# For Participants

- Introduce material for you to review in greater depth on your own
- I’ll direct you to further resources if you’d like to go beyond material covered in each week’s assignment
- Office hours: questions about assignments, your own projects
  - Chat with me on Teams to schedule

# Raise your hand if…

- You know what XML data looks like

# Raise your hand if…

- You know what XML data looks like
- You’ve worked with XML data

# Raise your hand if…

- You know what XML data looks like
- You’ve worked with XML data
- You’ve worked with ElementTree

# XML Data

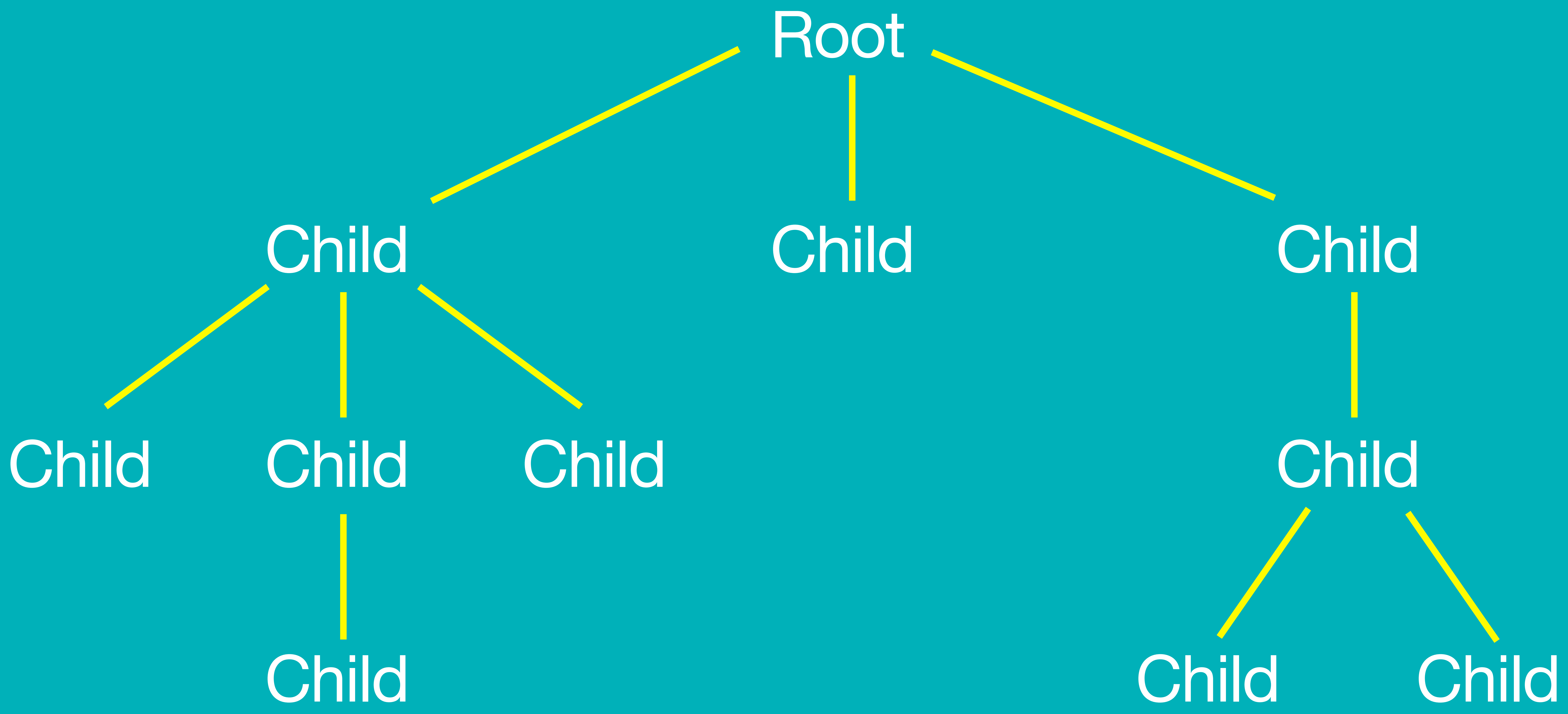
E**x**tensible **M**arkup **L**anguage

Similar to HTML (**H**yper**t**ext **M**arkup **L**anguage)

Hierarchical

Reference: https://www.datacamp.com/community/tutorials/python-xml-elementtree#intro

# XML Data



Root

Child Child Child

Child Child Child Child

Child Child Child

DEMO

# XML Example

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# XML Example

What’s the root?

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# XML Example

What's a child?

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# An XML Start-Tag

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# An XML Start-Tag

Where’s the end-tag?

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# XML Tags

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# An XML Element

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# An XML Attribute

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# An XML Attribute

Can you name others?

```
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

# XML Text

```xml
<?xml version="1.0"?>
<collection>
  <genre category="Action">
      <movie title="Indiana Jones: The raiders of the lost Ark">
        <format multiple="No">DVD</format>
        <year>1981</year>
        <rating>PG</rating>
        <description>
                'Archaeologist and adventurer Indiana Jones is hired
                by the U.S. government to find the Ark of the
                Covenant before the Nazis.'
        </description>
      </movie>
 </genre>
</collection>
```

DEMO

# XML in the Real World

- Web publishing
- Business applications (send data between different technical systems)
- Digital metadata formats
- Harvesting data through APIs
- Downloading data dumps

Reference: https://www.ibm.com/support/knowledgecenter/ssw_ibm_i_73/rzamj/rzamjintrouses.htm

# DEMO

# Assignment

Watch videos *1.2. Quick Overview of XML* and *6.3. The ElementTree API* in the course *Python: XML, JSON, and the Web*
https://www.linkedin.com/learning/python-xml-json-and-the-web/quick-overview-of-xml?u=50251009

**Find or create your own XML file to parse and analyze with ElementTree!** What questions can you ask about it using the methods and functions in ElementTree? Can you transform the XML into another data format and put it into a Pandas DataFrame?

# Thanks everyone!

Next course meeting: Friday, 10:00-11:00 AM BST

Office hours available on Wednesday (30 minutes)

*To schedule, please message me on Teams!*