NYU TANDON SCHOOL OF ENGINEERING

Lab 7
Websites

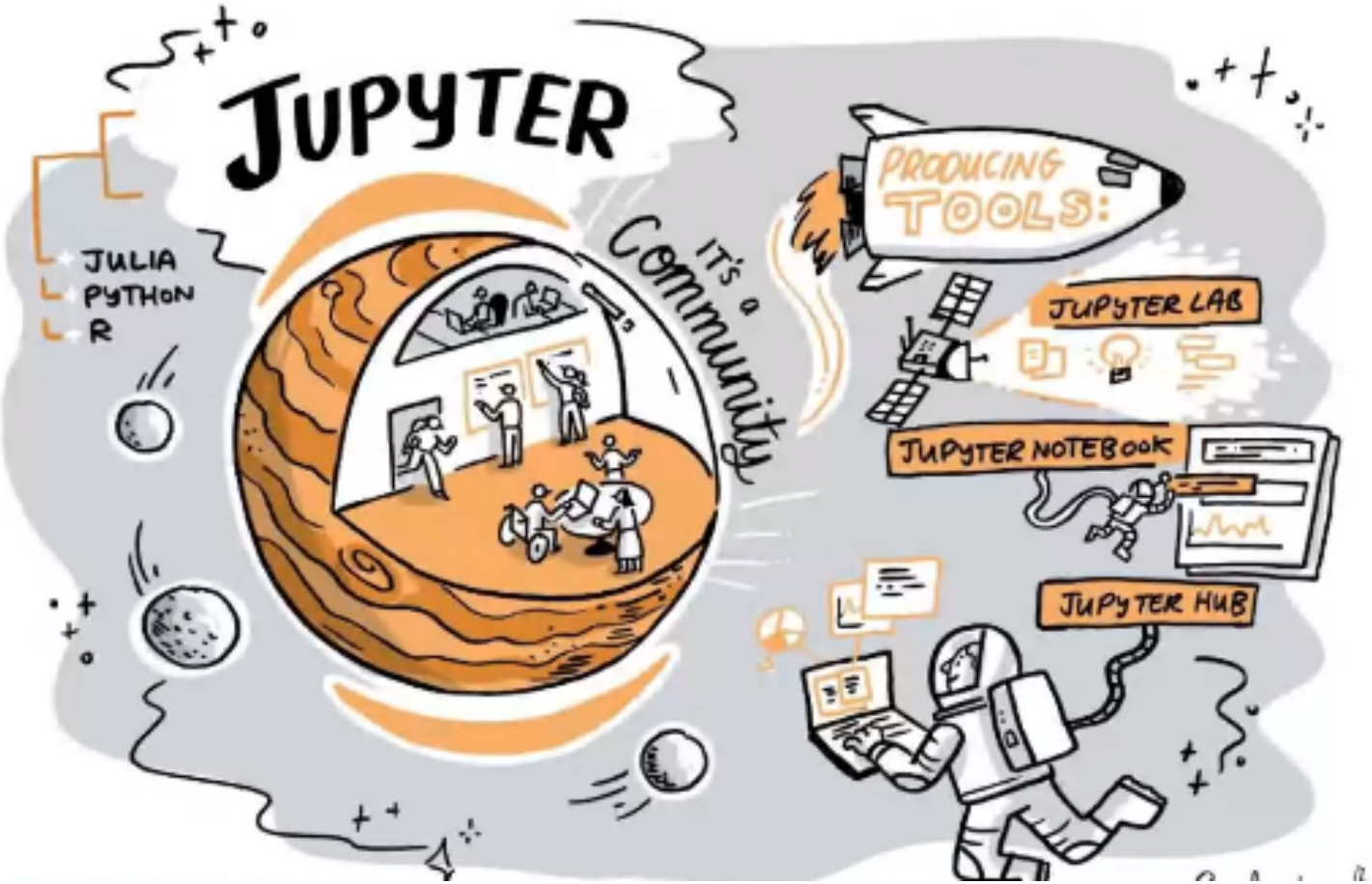
Programming for Business Analytics
MG-GY 8401

NYU TANDON SCHOOL OF ENGINEERING

Agenda

- Accessing Internet Data
- Parsing Data with XPath

Question & Answer



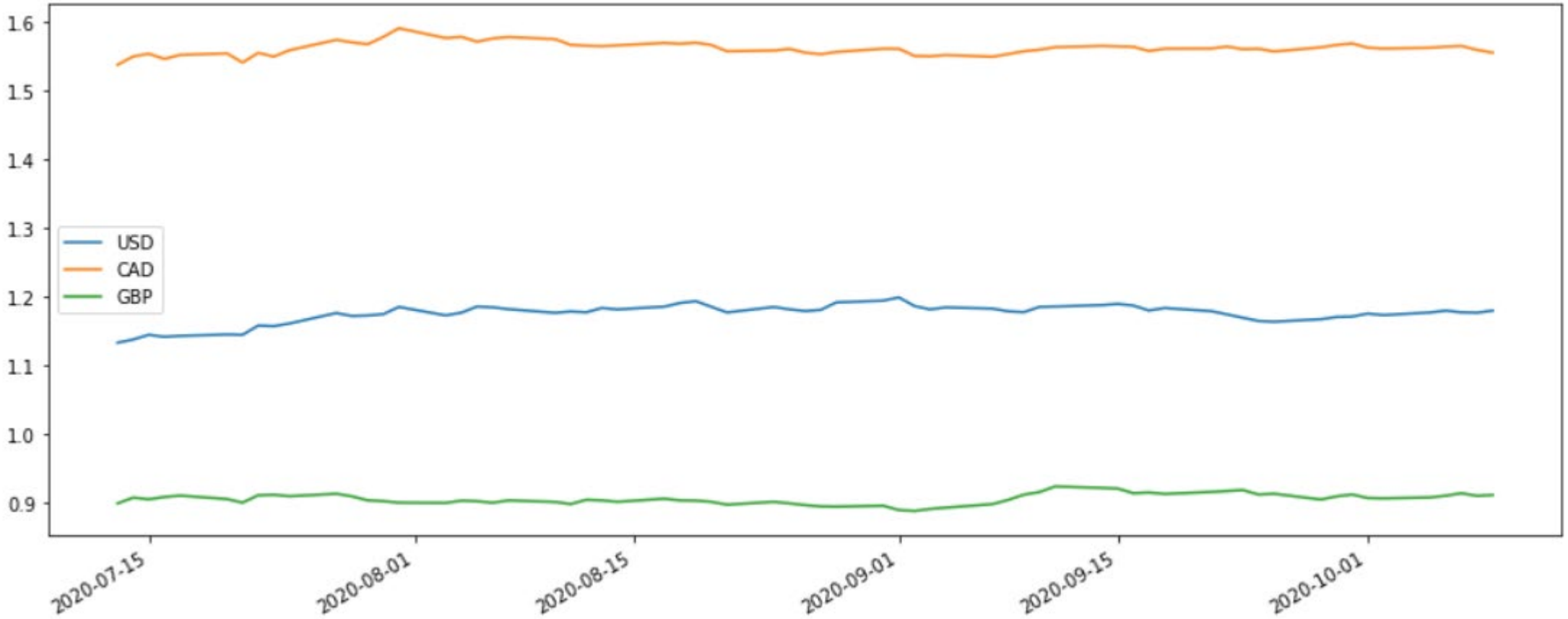
- Describe the learning objectives.
- Summarize the relevant take-aways.
- Ask about unclear information.

CC BY NC ND

The European Central Bank is the central bank of the European Union consisting of nineteen member countries using the euro.

CC BY NC ND

Case Study

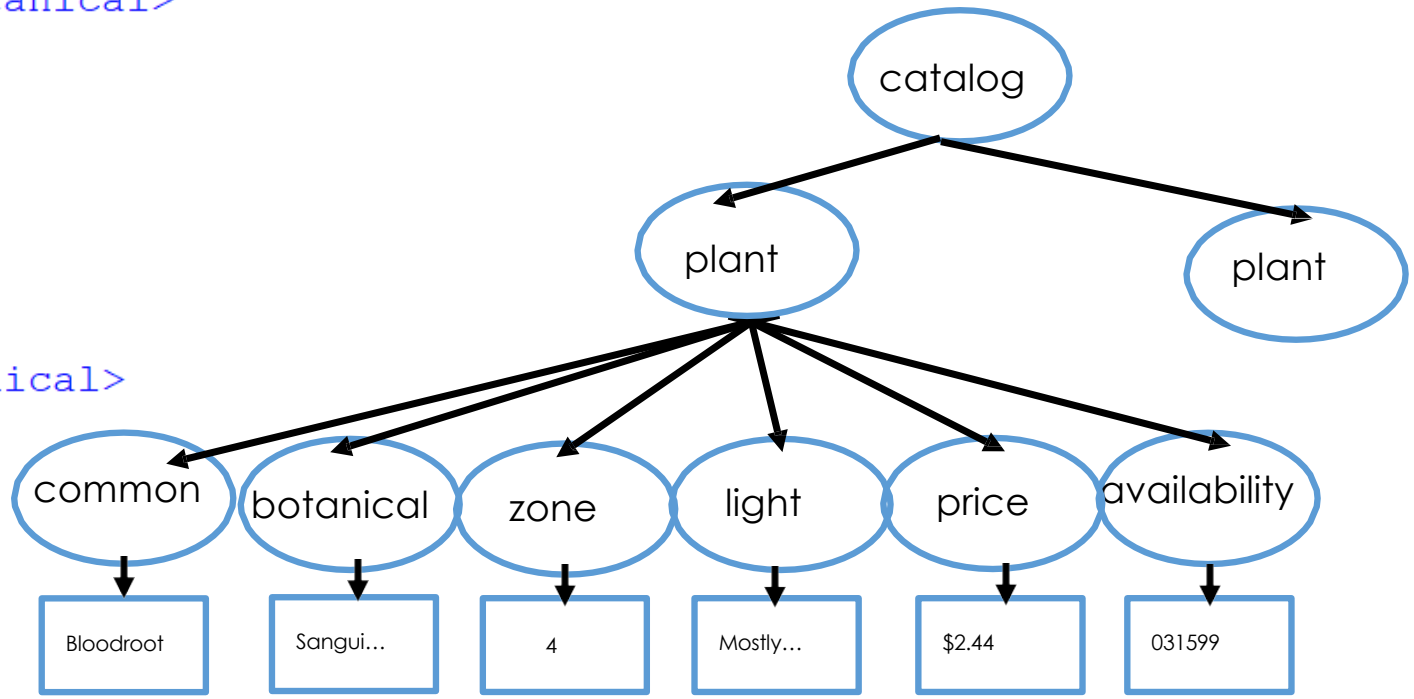


CC BY NC ND

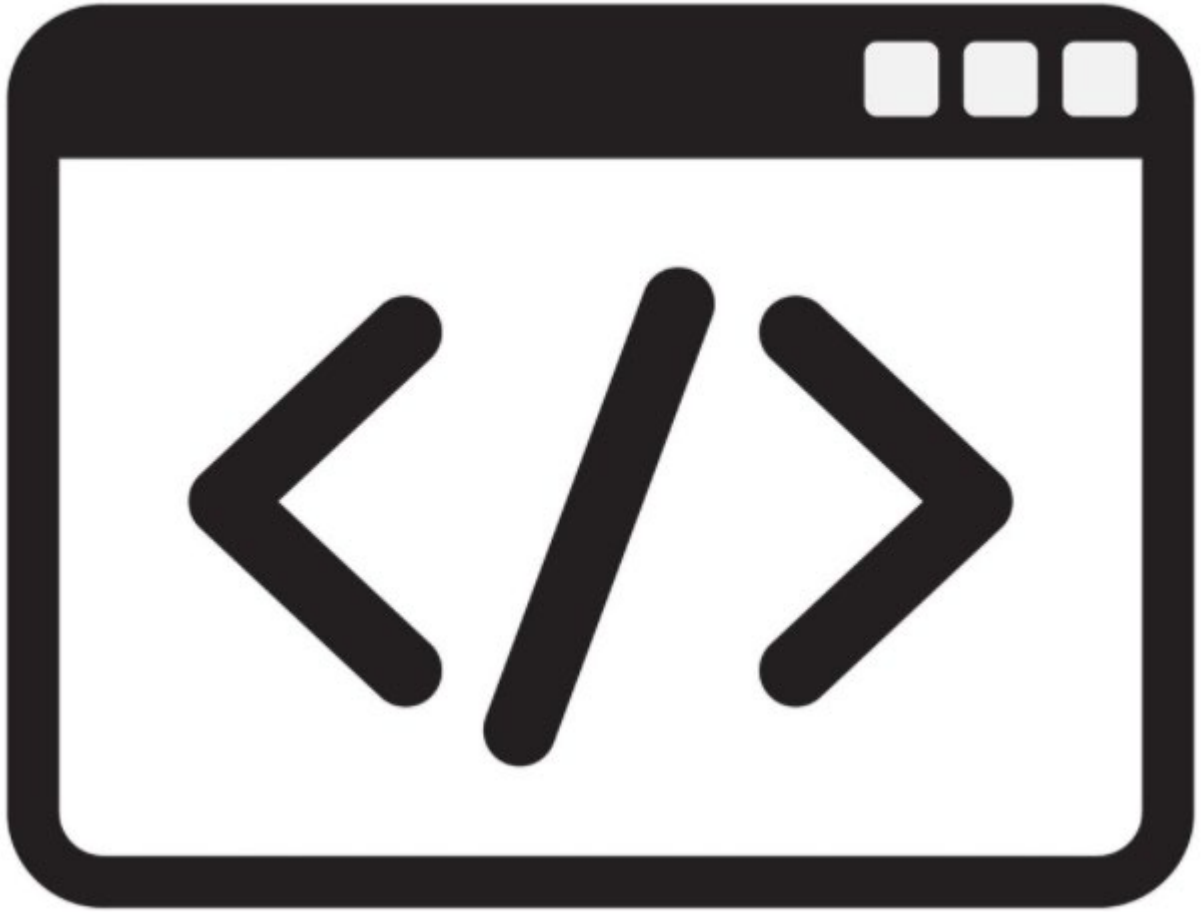
eXtensible Markup Language is a protocol for formatting data with a tree-like structure

CC BY NC ND

```
<catalog>
    <plant>
        <common>Bloodroot</common>
        <botanical>Sanguinaria canadensis</botanical>
        <zone>4</zone>
        <light>Mostly Shady</light>
        <price currency="USD">$2.44</price>
        <availability>031599</availability>
    </plant>
    <plant>
        <common>Columbine</common>
        <botanical>Aquilegia canadensis</botanical>
        <zone>3</zone>
        <light>Mostly Shady</light>
        <price currency="USD">$9.37</price>
        <availability>030699</availability>
    </plant>
    <plant>
        <common>Marsh Marigold</common>
        <botanical>Caltha palustris</botanical>
        <zone>4</zone>
        <light>Mostly Sunny</light>
        <price currency="CAD">$6.81</price>
        <availability>051799</availability>
    </plant>
</catalog>
```



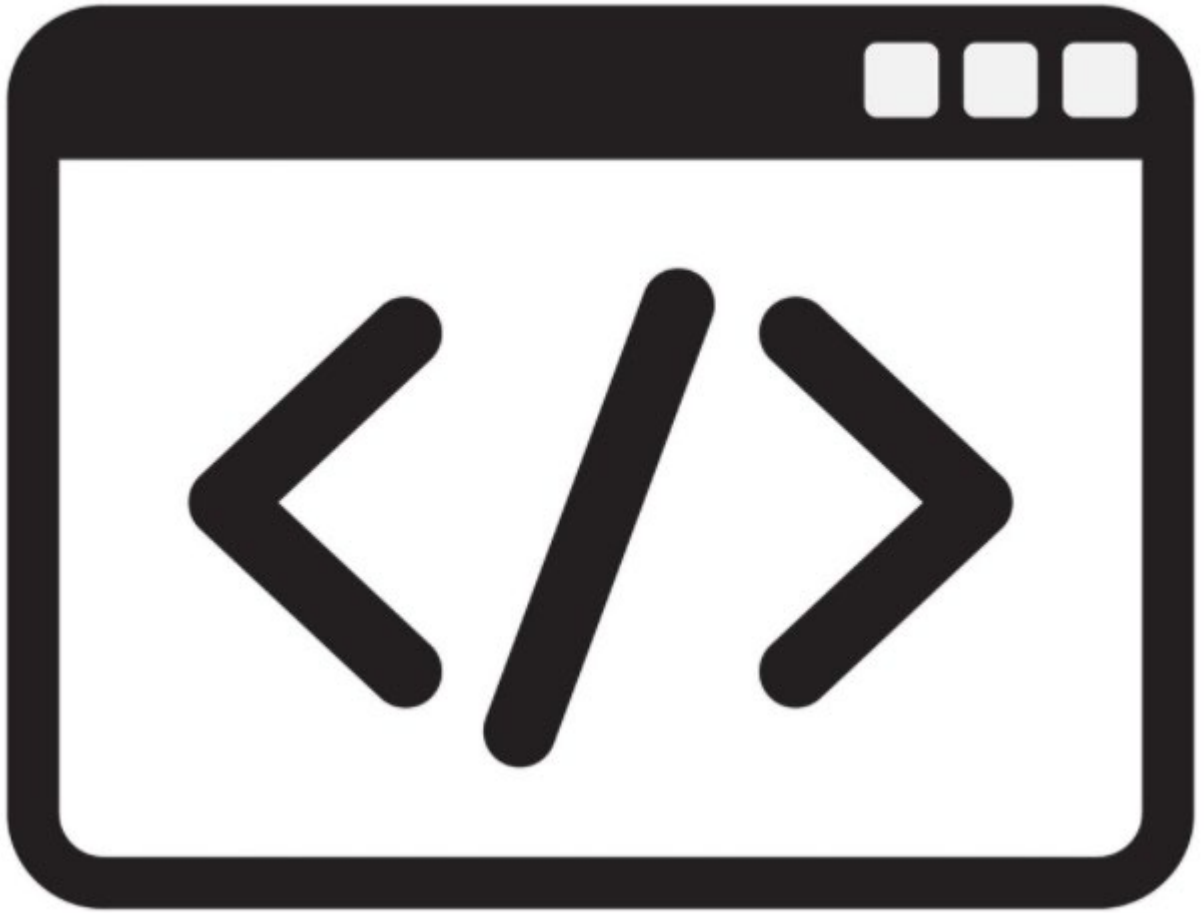
Extraction tool designed for locating content in an XML file or HTML file



CC BY NC ND

Expressions contain three components

- Direction
- Element Information
- Logical Expression



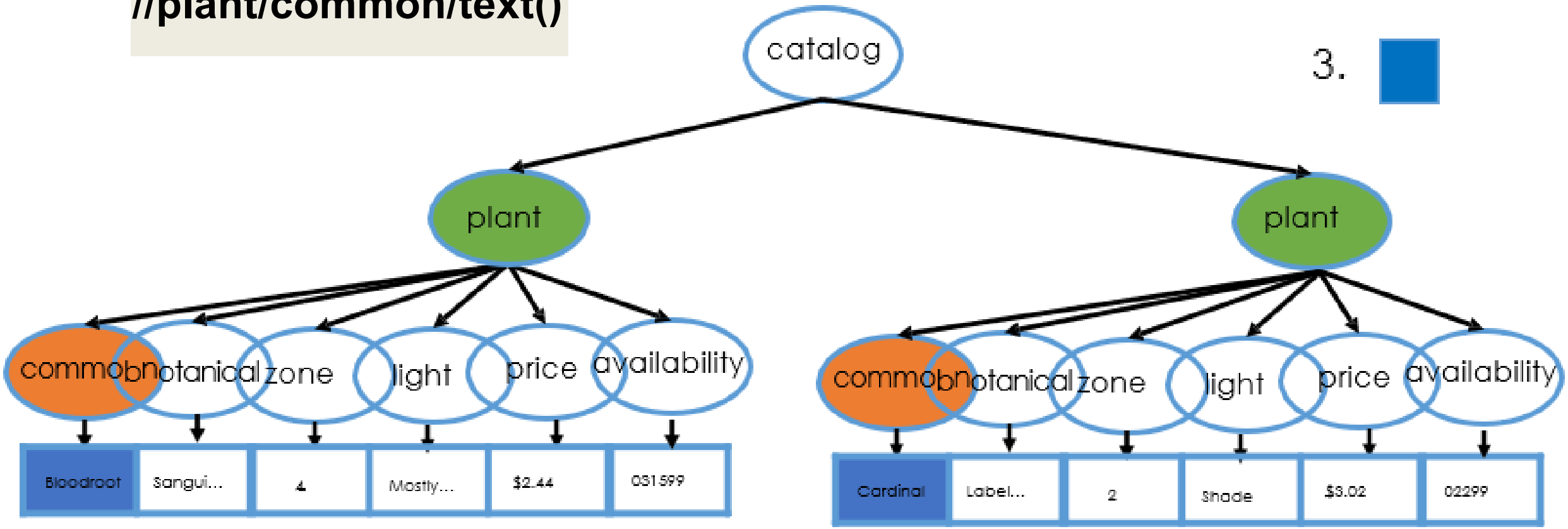
CC BY NC ND

Example

Retrieve common names of all plants

//plant/common/text()

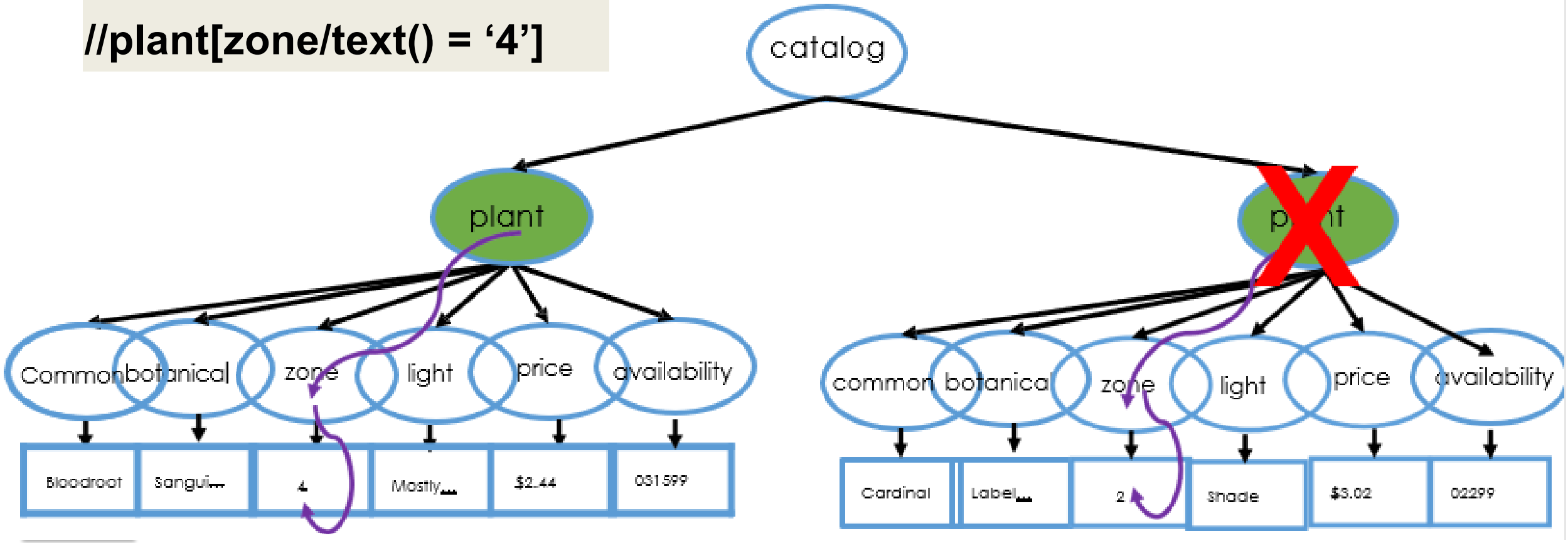
1.
2.
3.



Example

Retrieve plants that grow in zone 4

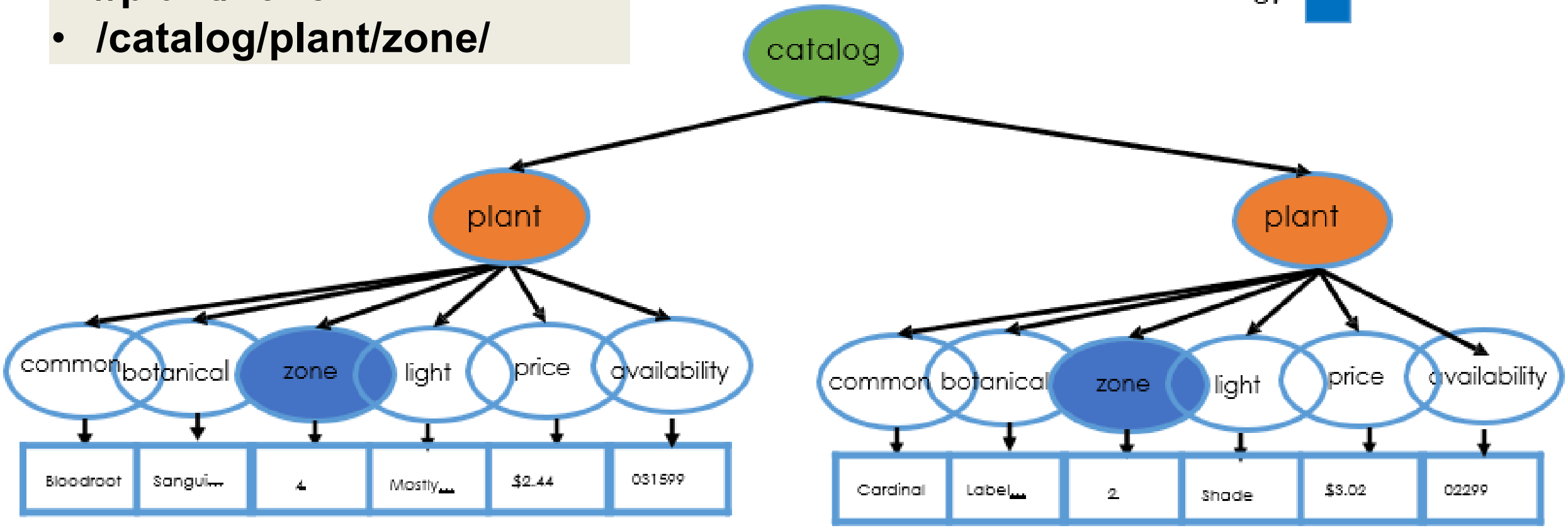
//plant[zone/text() = '4']

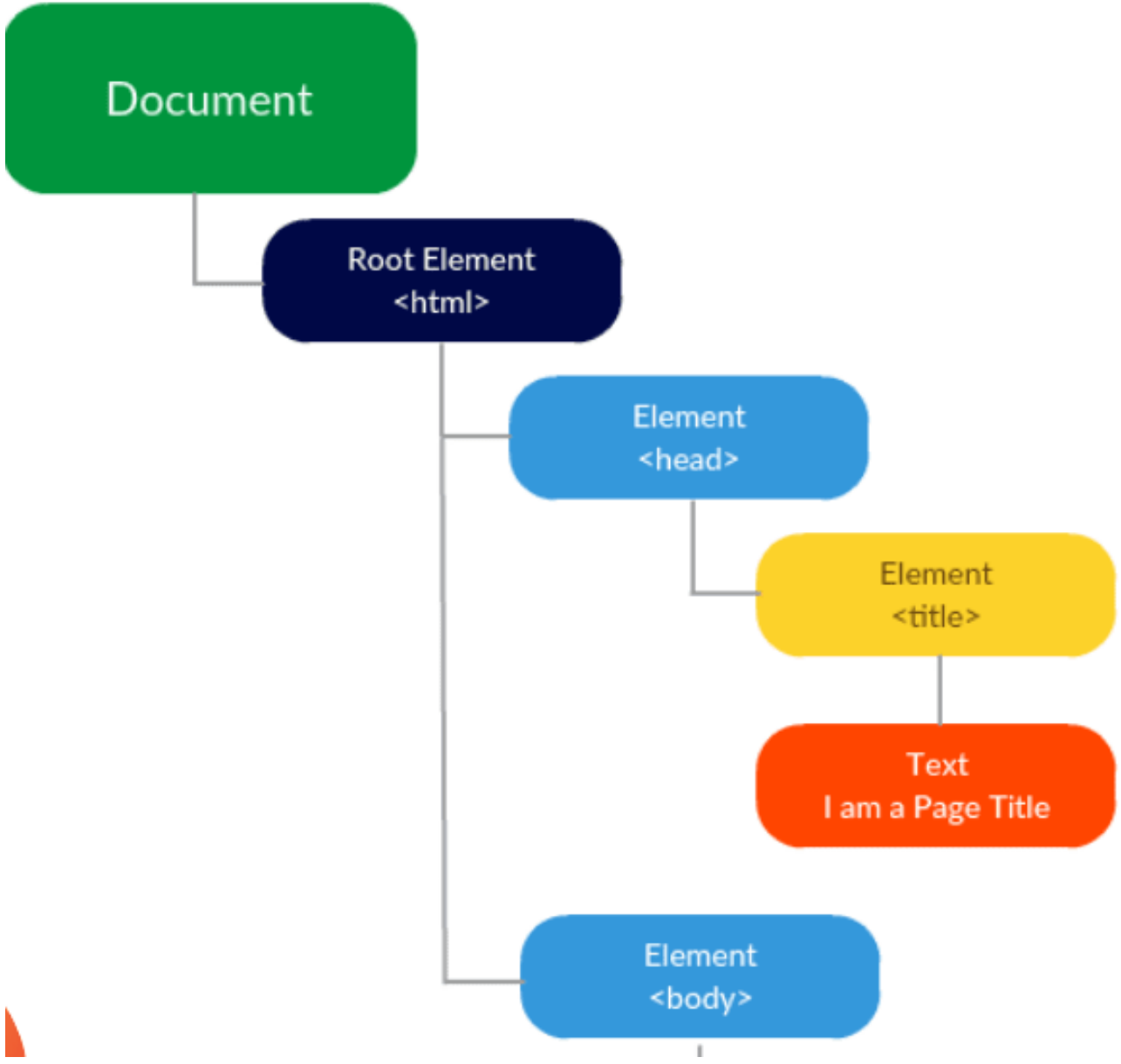


Three ways to access zone

- **//zone**
- **//plant/zone**
- **/catalog/plant/zone/**

1.
2.
3.





Directions

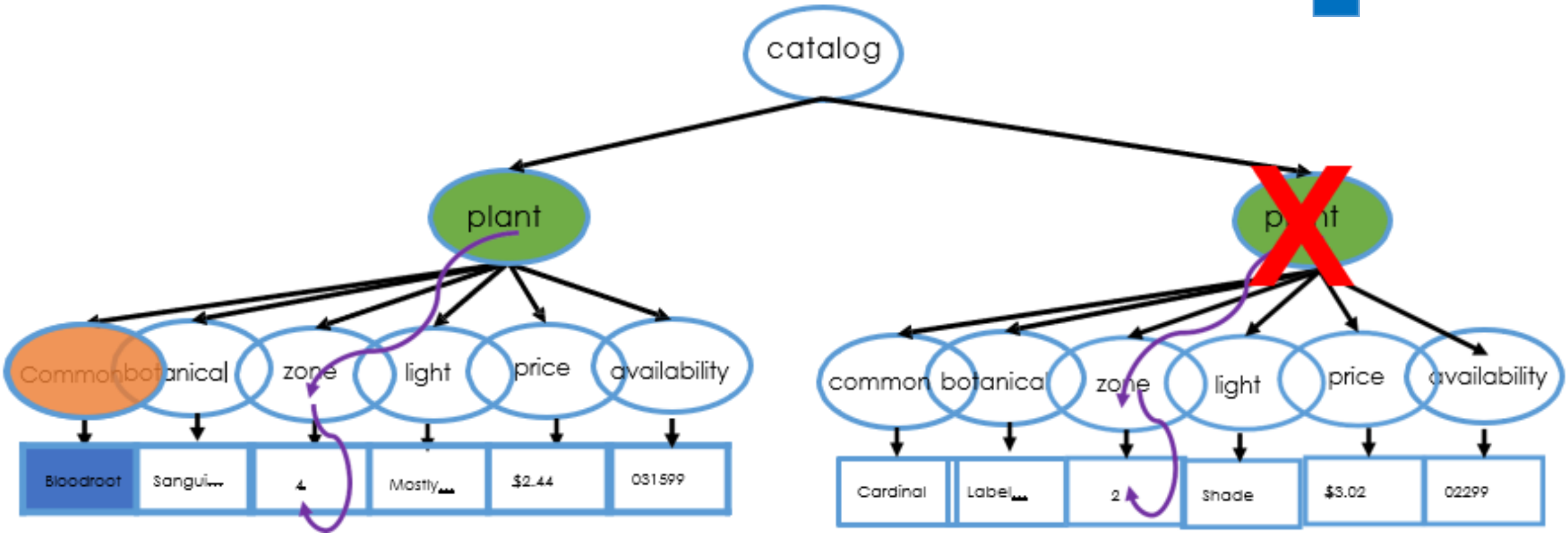
- Child /
- Descendent //
- Self .
- Parent ..

Example

Retrieve common names of plants that grow in zone 4

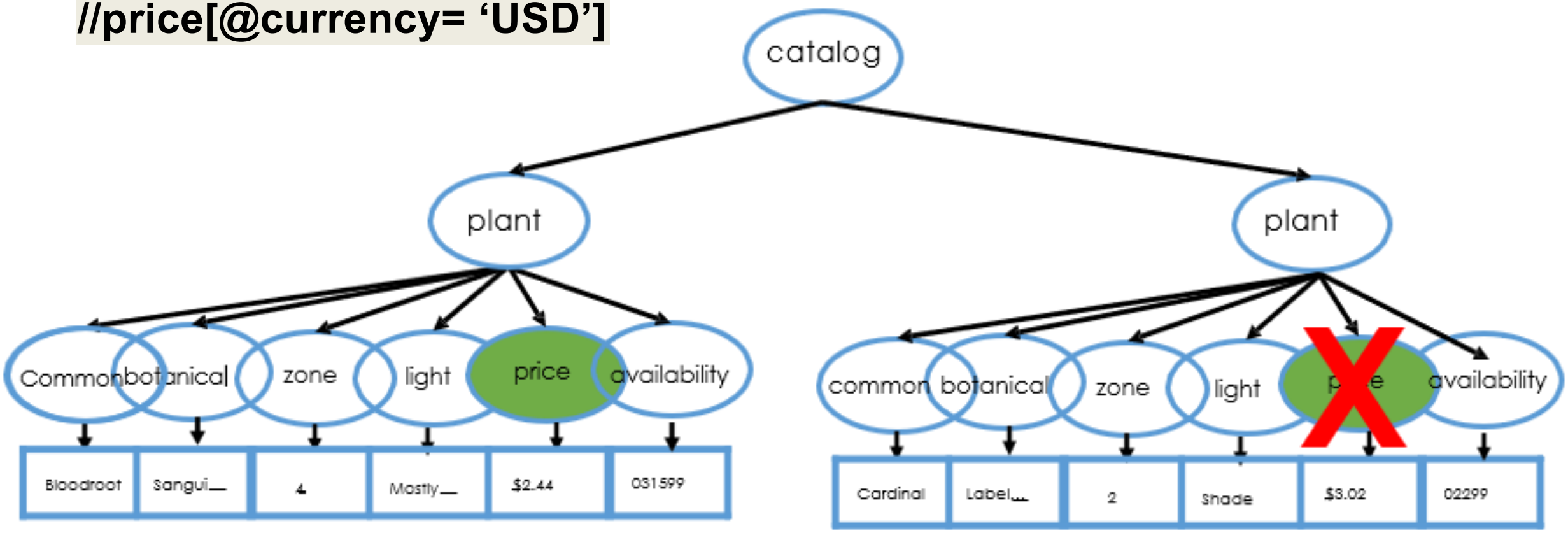
//common[../zone/text() = '4']/text()

1.
2.
3.

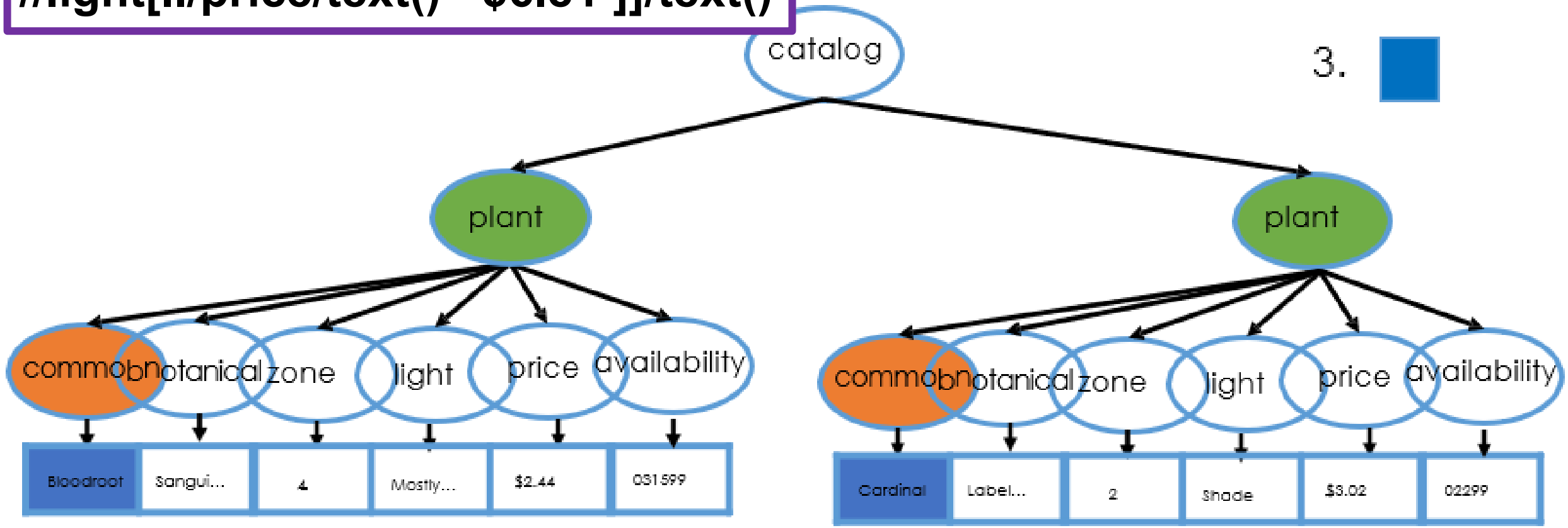


Retrieve prices of plants whose prices are listed in USD

//price[@currency= 'USD']



What is the result of the following Xpath query?

//light[../price/text()='$6.81']]/text()

1.
2.
3.