

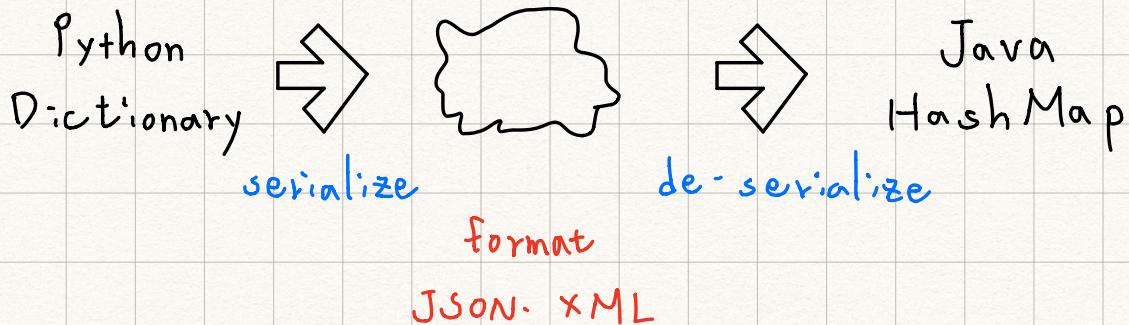
Data on the Web

Data on the Web

There was a natural move toward exchanging data between programs using these protocols

There are two commonly used formats:
XML and JSON

Sending Data across the "Net"



eXtensible Markup Language (XML)

XML "Elements" (or Nodes)

	<people>
	<person>
	<name> Chuck </name>
Simple Element	<phone> 303 4456 </phone>
	</person>
	<person>
	<name> Noah </name>
Complex Element	<phone> 622 7421 </phone>
	</person>
	</people>

eXtensible Markup Language (XML)

Primary purpose :
help information systems share structured data

It started as a simplified of the
HTML , and is designed to be relatively
human - legible

XML Basics

- Start Tag

< person >

- End Tag

< name > Chuck < /name >

- Text Content

< phone type = "intl" >

+1 734 303 4456

< /phone >

< email hide = "yes" />

< /person >

- Self Closing Tag

White Space

Line ends do not matter. White space is generally discarded on text elements. We indent only to be readable.

Some XML ...

<http://en.wikipedia.org/wiki/XML>

XML Terminology

Tags indicate the beginning and ending of elements

Attributes keyword/value pairs on the opening tag of XML

Serialize / De-Serialize convert data in one program into a common format that can be stored and / or transmitted between systems in a programming language - independent manner

XML as a Tree

<a>

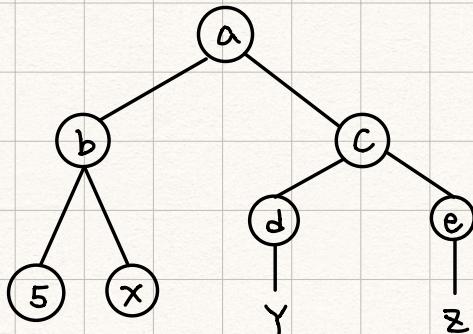
<b w="5"> X

<c>

<d> Y </d>

<e> Z </e>

</c>



/a/b - x

/a/c/e - z

XML Schema

XML Schema

Describing a "contract" as to what is acceptable XML.

- Description of the "legal format" of an XML document
- Often used to specify a "contract" between systems
- If a particular piece of XML meets the specification of the Schema — it is said to "validate"

XML Validation

XML Document

XML Schema Contract

Validator

Many XML Schema Languages

Document Type Definition (DTD)

Standard Generalized Markup Language

XML Schema from W3C - (XSD)

XSD Structure

xs : complexType xs : sequence xs : element

```
<xs:complexType name = "person">
  <xs:sequence>
    <xs:element name = ..... />
    <xs:element name = .... />
  </xs:sequence>
</xs:complexType>
```

http://www.w3schools.com/Schema/Schema_datatypes_numeric.asp

ISO 8601 Data / Time Format

2002-05-30T09:30:10Z

Year-month-day Time of day Timezone

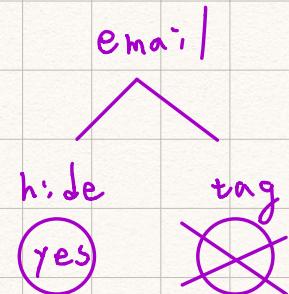
Parsing XML

import xml.etree.ElementTree as ET

```
tree = ET.fromstring(data)
print('Name:', tree.find('name').text)
print('Attr:', tree.find('email').get('hide'))
```

→ 將 data 轉成 tree
→ name → chuck
→ get 能取得 < > 裡的 `` 資料

```
data = '''<person>
<name>Chuck</name>
<phone type = "intl">
    +1 734 303 4456
</phone>
<email hide = "yes"/>
</person>'''
```



見上面的 data，若另一 XML data 中有許多這樣
的資料，並被 `<users>` 包起來，得先做
`</users>`

以下處理才能獲得 name, phone 的數據

```
stuff = ET.fromstring(input)
lst = stuff.findall('users/user')
for item in lst:
    print('Name', item.find('name').text)
```