***Python for everybody***

***Using Python to access web data***

***Week 5***

**Sending data across the “net”**

1- Agreeing on a Wire Format

* the act of going from an internal representation on one computer out to a sort of interchange format is called ***serialization***
* So it was taking, from the internal memory of the computer a format that we could sort of sending one character at a time, character, character, character, character, character, so we called this a ***serialization format***
* the act of taking the data off of the wire and turning it into a new internal data structure, in the new environment, potentially in a very new language, is called ***de-serialization***
* you take your Python dictionary, you produce JSON. You send JSON across the network as a string or a document, and then you receive the document, and then you turn it into whatever it is it's going to be on that far side. So that's the basic idea of ***agreeing on data formats***

2- eXtensible Markup Language (XML)

* **tags** are the beginning and the end.
* **Attributes** are key-value pairs on the start tag.
* **Serialization** is the act of taking from an internal structure in one programming environment, sending it across the network.
* **Deserialization** is received across the network and translated it back into an internal structure on the destination computer.

3- XML Scheme

* define a contract as to what is acceptable XML
* It's a way to sort of establishing outside of any program, and then separately check
* **Validation** is the act of verifying that the data is in the right format
* **XML validation** is the act of taking a document and a Schema Contract, which itself is also an XML document, and then sending to the Validator.

4- Parsing XML

* importing a library xml.etree.ElementTree