Using Web Services Chapter 13

Python for Informatics: Exploring Information www.py4inf.com

open.michigan

Unless otherwise noted, the content of this course material is licensed under a Creative Commons Attribution 3.0 License.

http://creativecommons.org/licenses/by/3.0/.

Copyright 2009- Charles Severance.







Data on the Web

- With the HTTP Request/Response well understood and well supported there was a natural move toward exchanging data between programs using these protocols
- We needed to come up with an agreed way to represent data going between applications and across networks
- There are two commonly used formats: XML and JSON

Sending Data across the "Net"



a.k.a. "Wire Protocol" - What we send on the "wire"

Agreeing on a "Wire Format"

<person> <name> De-Serialize Chuck Python </name> Java Dictionary HashMap <phone> 303 4456 Serialize </phone> </person> **XML**

Agreeing on a "Wire Format"

```
Python
Dictionary

| Chuck", "phone": "303-4456" | HashMap | HashMap | Phone |
```

XML "Elements" (or Nodes)

- Simple Element
- Complex Element

```
<people>
 <person>
  <name>Chuck</name>
  <phone>303 4456</phone>
 </person>
 <person>
  <name>Noah</name>
  <phone>622 7421</phone>
 </person>
</people>
```

XML

Marking up data to send across the network...

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

eXtensible Markup Language

- Primary purpose is to help information systems share structured data
- It started as a simplified subset of the Standard Generalized Markup Language (SGML), and is designed to be relatively human-legible

XML Basics

- Start Tag
- End Tag
- Text Content
- Attribute
- Self Closing Tag

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

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

White Space

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

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

Some XML...

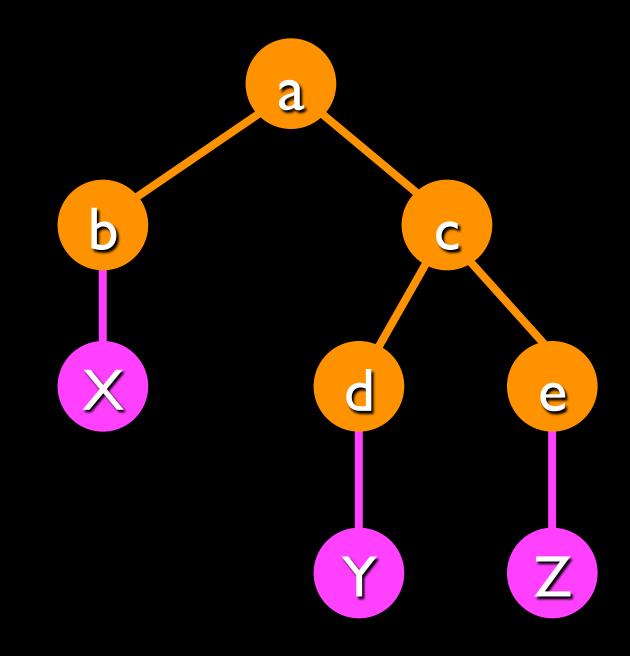
```
<recipe name="bread" prep_time="5 mins" cook_time="3 hours">
 <title>Basic bread</title>
  <ingredient amount="8" unit="dL">Flour</ingredient>
  <ingredient amount="10" unit="grams">Yeast</ingredient>
  <ingredient amount="4" unit="dL" state="warm">Water</ingredient>
  <ingredient amount="1" unit="teaspoon">Salt</ingredient>
  <instructions>
    <step>Mix all ingredients together.</step>
    <step>Knead thoroughly.</step>
    <step>Cover with a cloth, and leave for one hour in warm room.</step>
    <step>Knead again.</step>
    <step>Place in a bread baking tin.</step>
    <step>Cover with a cloth, and leave for one hour in warm room.</step>
    <step>Bake in the oven at 180(degrees)C for 30 minutes.
 </instructions>
</recipe>
```

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

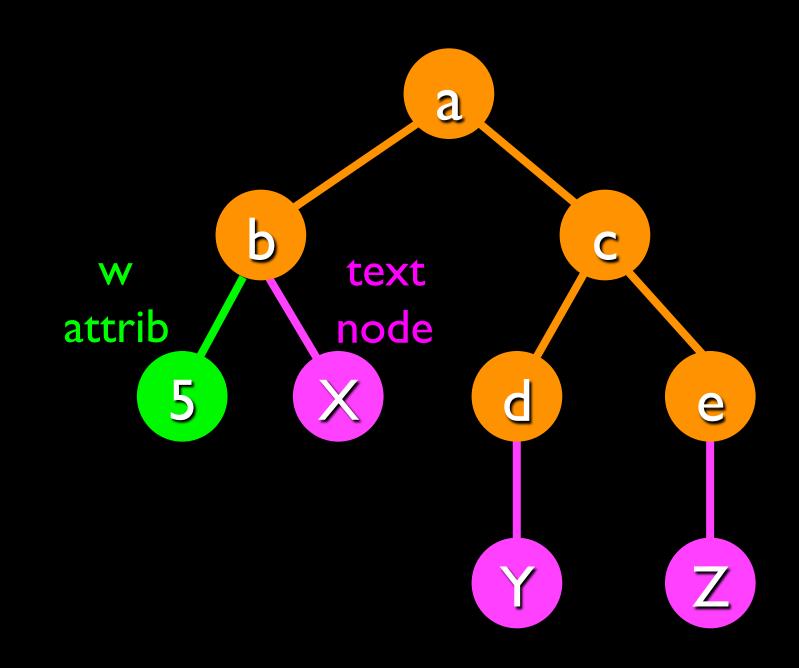
XML as a Tree

```
<a>
<b>X</b>
<c>
<d>Y</d>
<e>Z</e>
</c>
</a>
```

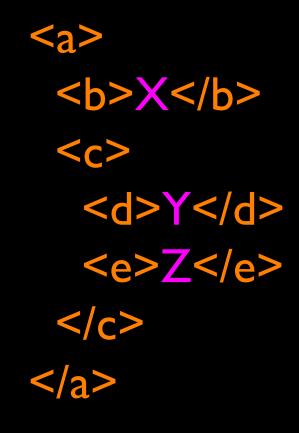


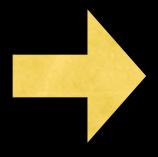
XML Text and Attributes

```
<a>
<b w="5">X</b>
<b
```



XML as Paths

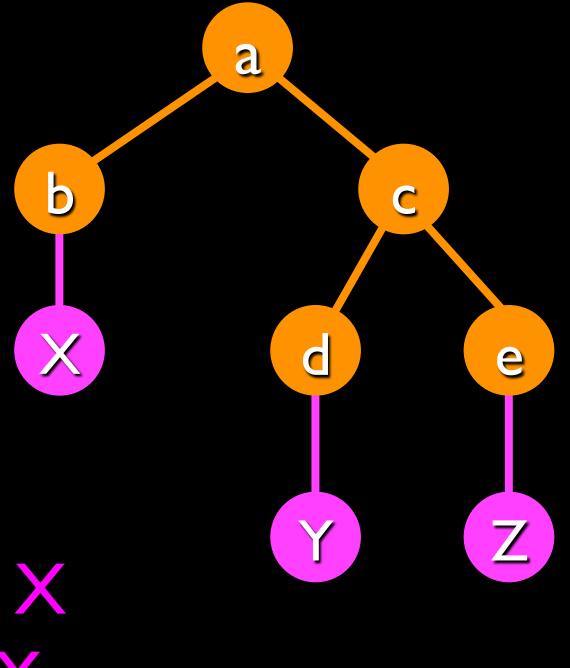




/a/b /a/c/d /a/c/e







XML Schema

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

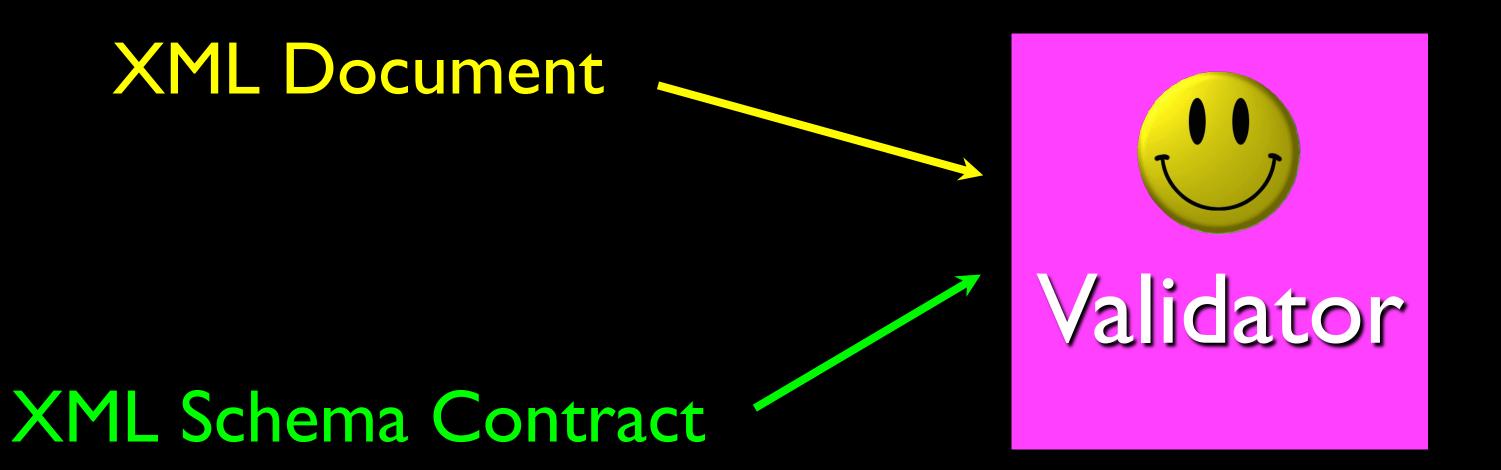
http://en.wikipedia.org/wiki/Xml_schema http://en.wikibooks.org/wiki/XML_Schema

XML Schema

- Description of the legal format of an XML document
- Expressed in terms of constraints on the structure and content of documents
- Often used to specify a "contract" between systems "My system will only accept XML that conforms to this particular Schema."
- If a particular piece of XML meets the specification of the Schema it is said to "validate"

http://en.wikipedia.org/wiki/Xml_schema

XML Validation



XML Document

</xs:complexType>

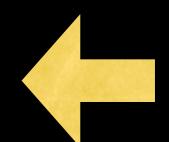
XML Validation

```
<person>
     <lastname>Severance/lastname>
     <age>|7</age>
     <dateborn>2001-04-17</dateborn>
   </person>
     XML Schema Contract
<xs:complexType name="person">
 <xs:sequence>
  <xs:element name="lastname" type="xs:string"/>
  <xs:element name="age" type="xs:integer"/>
  <xs:element name="dateborn" type="xs:date"/>
 </xs:sequence>
```



Many XML Schema Languages

- Document Type Definition (DTD)
 - http://en.wikipedia.org/wiki/Document_Type_Definition
- Standard Generalized Markup Language (ISO 8879:1986 SGML)
 - http://en.wikipedia.org/wiki/SGML
- XML Schema from W3C (XSD)
 - http://en.wikipedia.org/wiki/XML_Schema_(W3C)



XSD XML Schema (W3C spec)

- We will focus on the World Wide Web Consortium (W3C) version
- It is often called "W3C Schema" because "Schema" is considered generic
- More commonly it is called XSD because the file names end in .xsd

http://www.w3.org/XML/Schema

http://en.wikipedia.org/wiki/XML_Schema_(W3C)

XSD Structure

- xs:element
- xs:sequence
- xs:complexType

```
<person>
    <lastname>Severance</lastname>
        <age>17</age>
        <dateborn>2001-04-17</dateborn>
</person>
```

```
<xs:complexType name="person">
  <xs:sequence>
    <xs:element name="lastname" type="xs:string"/>
    <xs:element name="age" type="xs:integer"/>
    <xs:element name="dateborn" type="xs:date"/>
    </xs:sequence>
</xs:complexType>
```

```
<xs:element name="person">
                                                           XSD
 <xs:complexType>
  <xs:sequence>
                                                    Constraints
   <xs:element name="full name" type="xs:string"</pre>
minOccurs="I" maxOccurs="I" />
   <xs:element name="child_name" type="xs:string"</pre>
minOccurs="0" maxOccurs="10" />
  </xs:sequence>
                                       <person>
 </xs:complexType>
                                        <full name>Tove Refsnes</full name>
</xs:element>
                                        <child name>Hege</child name>
                                        <child name>Stale</child name>
                                        <child_name>Jim</child_name>
                                        <child name>Borge</child name>
                                       </person>
```

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

XSD Data Types

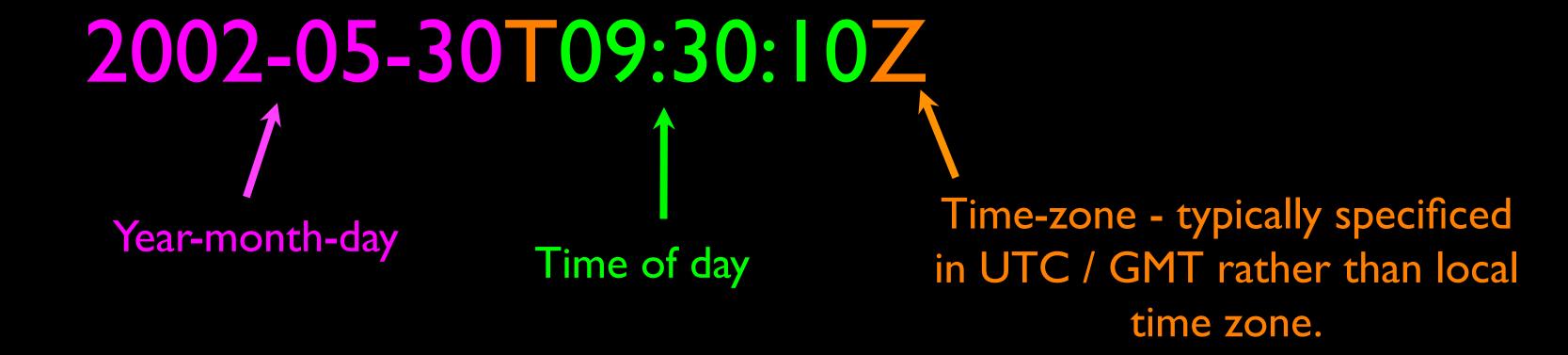
```
<xs:element name="customer" type="xs:string"/>
<xs:element name="start" type="xs:date"/>
<xs:element name="startdate" type="xs:dateTime"/>
<xs:element name="prize" type="xs:decimal"/>
<xs:element name="weeks" type="xs:integer"/>
```

```
It is common to represent time in UTC/GMT given that servers are often scattered around the world.
```

```
<customer>John Smith</customer>
<start>2002-09-24</start>
<startdate>2002-05-30T09:30:10Z</startdate>
<pri><pri><pri><pri>ze>999.50</pri>
<weeks>30</weeks>
```

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

ISO 8601 Date/Time Format



http://en.wikipedia.org/wiki/ISO_8601

http://en.wikipedia.org/wiki/Coordinated_Universal_Time

```
<?xml version="1.0" encoding="utf-8" ?>
<xs:schema elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Address">
   <xs:complexType>
     <xs:sequence>
        <xs:element name="Recipient" type="xs:string" />
        <xs:element name="House" type="xs:string" />
        <xs:element name="Street" type="xs:string" />
        <xs:element name="Town" type="xs:string" />
        <xs:element minOccurs="0" name="County" type="xs:string" />
        <xs:element name="PostCode" type="xs:string" />
        <xs:element name="Country">
         <xs:simpleType>
            <xs:restriction base="xs:string">
             <xs:enumeration value="FR" />
             <xs:enumeration value="DE" />
             <xs:enumeration value="ES" />
                                              <?xml version="1.0" encoding="utf-8"?>
             <xs:enumeration value="UK" />
                                              <Address
             <xs:enumeration value="US" />
                                                 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance'
            </xs:restriction>
                                                 xsi:noNamespaceSchemaLocation="SimpleAddress.xsd">
         </xs:simpleType>
                                                <Recipient>Mr. Walter C. Brown/Recipient>
        </xs:element>
                                                <House>49</House>
     </xs:sequence>
                                                <Street>Featherstone Street/Street
   </r></xs:complexType>
                                                <Town>LONDON</Town>
 </xs:element>
                                                <PostCode>EC1Y 8SY</PostCode>
</xs:schema>
                                                <Country>UK</Country>
                                              </Address>
```

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="shiporder">
<xs:complexType>
  <xs:sequence>
   <xs:element name="orderperson" type="xs:string"/>
   <xs:element name="shipto">
    <xs:complexType>
     <xs:sequence>
      <xs:element name="name" type="xs:string"/>
      <xs:element name="address" type="xs:string"/>
      <xs:element name="city" type="xs:string"/>
      <xs:element name="country" type="xs:string"/>
     </xs:sequence>
    </xs:complexType>
   </xs:element>
   <xs:element name="item" maxOccurs="unbounded">
    <xs:complexType>
     <xs:sequence>
      <xs:element name="title" type="xs:string"/>
      <xs:element name="note" type="xs:string" minOccurs="0"/>
      <xs:element name="quantity" type="xs:positiveInteger"/>
      <xs:element name="price" type="xs:decimal"/>
     </xs:sequence>
    </xs:complexType>
   </xs:element>
 </xs:sequence>
  <xs:attribute name="orderid" type="xs:string" use="required"/>
</xs:complexType>
</xs:element>
</xs:schema>
```

```
import xml.etree.ElementTree as ET
data = '''<person>
  <name>Chuck</name>
  <phone type="intl">
     +1 734 303 4456
   </phone>
   <email hide="yes"/>
</person>''
tree = ET.fromstring(data)
print 'Name:',tree.find('name').text
print 'Attr:',tree.find('email').get('hide')
```

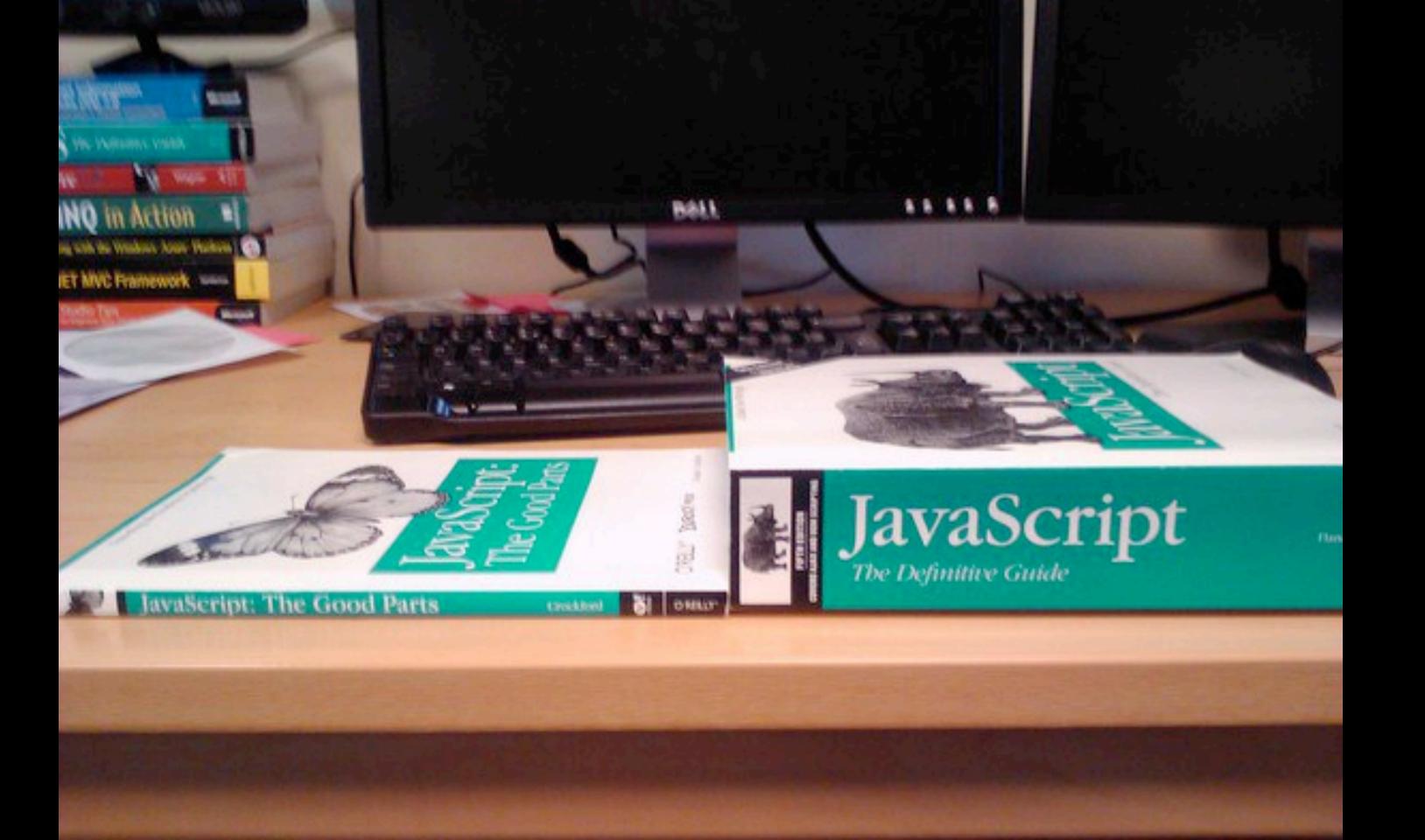
```
import xml.etree.ElementTree as ET
input = '''<stuff>
    <users>
        <user x="2">
            <id>001</id>
            <name>Chuck</name>
        </user>
        <user x="7">
            <id>009</id>
            <name>Brent</name>
        </user>
    </users>
</stuff>'''
stuff = ET.fromstring(input)
lst = stuff.findall('users/user')
print 'User count:', len(lst)
for item in 1st:
    print 'Name', item.find('name').text
    print 'Id', item.find('id').text
    print 'Attribute', item.get("x")
```

JavaScript Object Notation

JavaScript Object Notation

- Douglas Crockford "Discovered" JSON
- Object literal notation in JavaScript







Introducing JSON

Български 中文 Český Nederlandse Dansk English Esperanto Française Deutsch Ελληνικά עברית Magyar Indonesia Italiano 日本 한국어 فارسى Polski Português Română Русский Српски Slovenščina Español Svenska Türkçe Tiếng Việt

JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.

JSON is built on two structures:

- A collection of name/value pairs. In various languages, this is realized as an object, record, struct, dictionary, hash table, keyed list, or associative array.
- An ordered list of values. In most languages, this is realized as an array, vector, list, or sequence.

These are universal data structures. Virtually all modern programming languages support them in one form or another. It makes sense that a data format that is interchangeable with programming languages also be based on these structures.

In JSON, they take on these forms:

An object is an unordered set of name/value pairs. An object begins with { (left brace) and ends with } (right

```
object
      { members }
members
      pair
      pair, members
pair
      string: value
array
      [ elements ]
elements
      value
      value, elements
value
      string
      number
      object
```

```
import json
data = '''{
  "name" : "Chuck",
  "phone" : {
   "type" : "intl",
    "number" : "+1 734 303 4456"
   },
   "email" : {
    "hide" : "yes"
info = json.loads(data)
print 'Name:',info["name"]
print 'Hide:',info["email"]["hide"]
```

JSON represents data as nested "lists" and "dictionaries"

json2.py

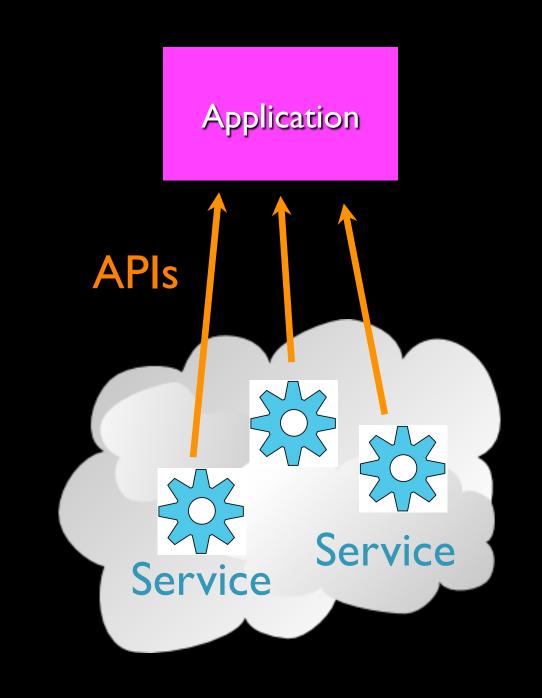
```
import json
input = '''[
 { "id" : "001",
   "x" : "2",
   "name" : "Chuck"
  { "id" : "009",
   "x" : "7",
   "name" : "Chuck"
info = json.loads(input)
print 'User count:', len(info)
for item in info:
    print 'Name', item['name']
    print 'Id', item['id']
    print 'Attribute', item['x']
```

JSON represents data as nested "lists" and "dictionaries"

Service Oriented Approach

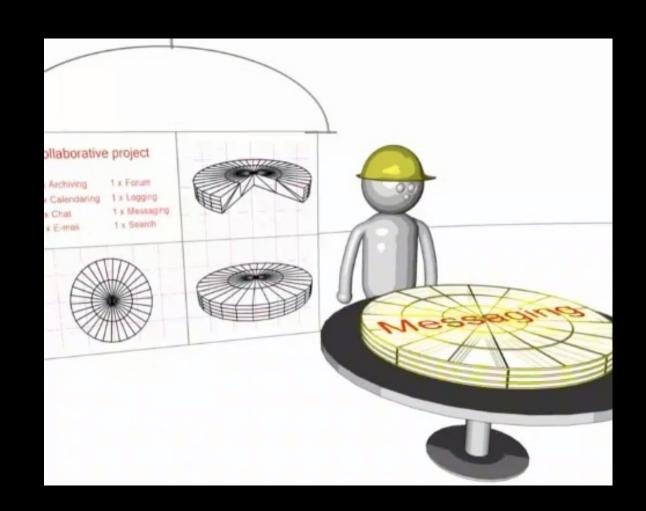
Service Oriented Approach

- Most non-trivial web applications use services
- They use services from other applications
 - Credit Card Charge
 - Hotel Reservation systems
- Services publish the "rules" applications must follow to make use of the service (API)



Multiple Systems

- Initially two systems cooperate and split the problem
- As the data/service becomes useful multiple applications want to use the information / application



Web Services

http://en.wikipedia.org/wiki/Web_services

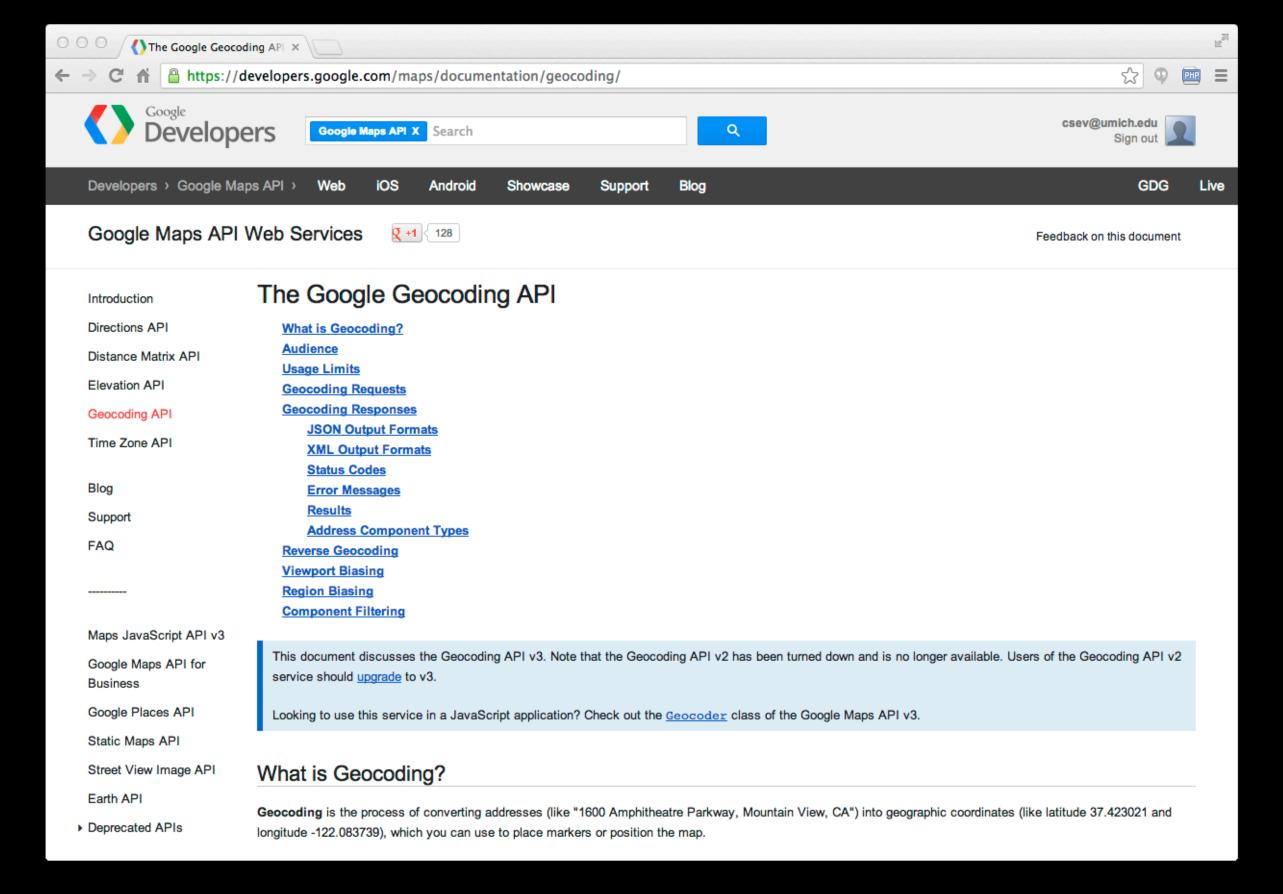
Application Program Interface

The API itself is largely abstract in that it specifies an interface and controls the behavior of the objects specified in that interface. The software that provides the functionality described by an API is said to be an "implementation" of the API. An API is typically defined in terms of the programming language used to build an application.

Web Service Technologies

- SOAP Simple Object Access Protocol (software)
 - Remote programs/code which we use over the network
 - Note: Dr. Chuck does not like SOAP because it is overly complex
- REST Representational State Transfer (resource focused)
 - Remote resources which we create, read, update and delete remotely

http://en.wikipedia.org/wiki/SOAP_(protocol) http://en.wikipedia.org/wiki/REST



```
"status": "OK",
"results": [
        "geometry": {
            "location type": "APPROXIMATE",
             "location": {
                "lat": 42.2808256,
                 "lng": -83.7430378
        "address components": [
                "long name": "Ann Arbor",
                 "types": [
                    "locality",
                     "political"
                "short name": "Ann Arbor"
         "formatted address": "Ann Arbor, MI, USA",
         "types": [
            "locality",
            "political"
```

http://maps.googleapis.com/maps/api/geocode/json?sensor=false&address=Ann+Arbor%2C+Ml

```
import urllib
import json
serviceurl = 'http://maps.googleapis.com/maps/api/geocode/json?' Retrieving http://maps.googleapis.com/...
while True:
    address = raw input('Enter location: ')
    if len(address) < 1 : break
    url = serviceurl + urllib.urlencode({'sensor':'false',
           'address': address})
    print 'Retrieving', url
    uh = urllib.urlopen(url)
    data = uh.read()
    print 'Retrieved',len(data),'characters'
    try: js = json.loads(str(data))
    except: js = None
    if 'status' not in js or js['status'] != 'OK':
        print '==== Failure To Retrieve ===='
        print data
        continue
    print json.dumps(js, indent=4)
    lat = js["results"][0]["geometry"]["location"]["lat"]
    lng = js["results"][0]["geometry"]["location"]["lng"]
    print 'lat', lat, 'lng', lng
    location = js['results'][0]['formatted address']
    print location
```

Enter location: Ann Arbor, MI Retrieved 1669 characters lat 42.2808256 lng -83.7430378 Ann Arbor, MI, USA Enter location:

API Security and Rate Limiting

- The compute resources to run these APIs are not "free"
- The data provided by these APIs is usually valuable
- The data providers might limit the number of requests per day, demand an API "key" or even charge for usage
- They might change the rules as things progress...

Usage Limits

The Google Geocoding API has the following limits in place:

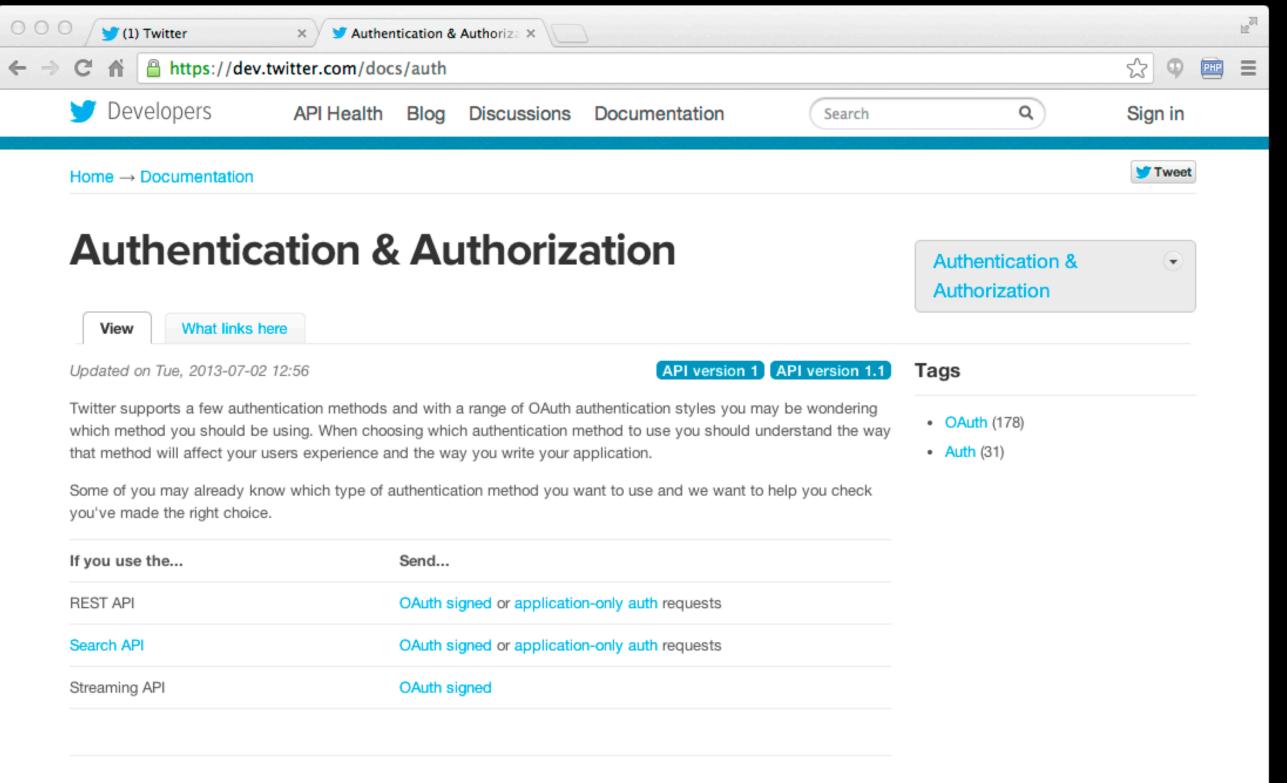
2,500 requests per day.

Google Maps API for Business customers have higher limits:

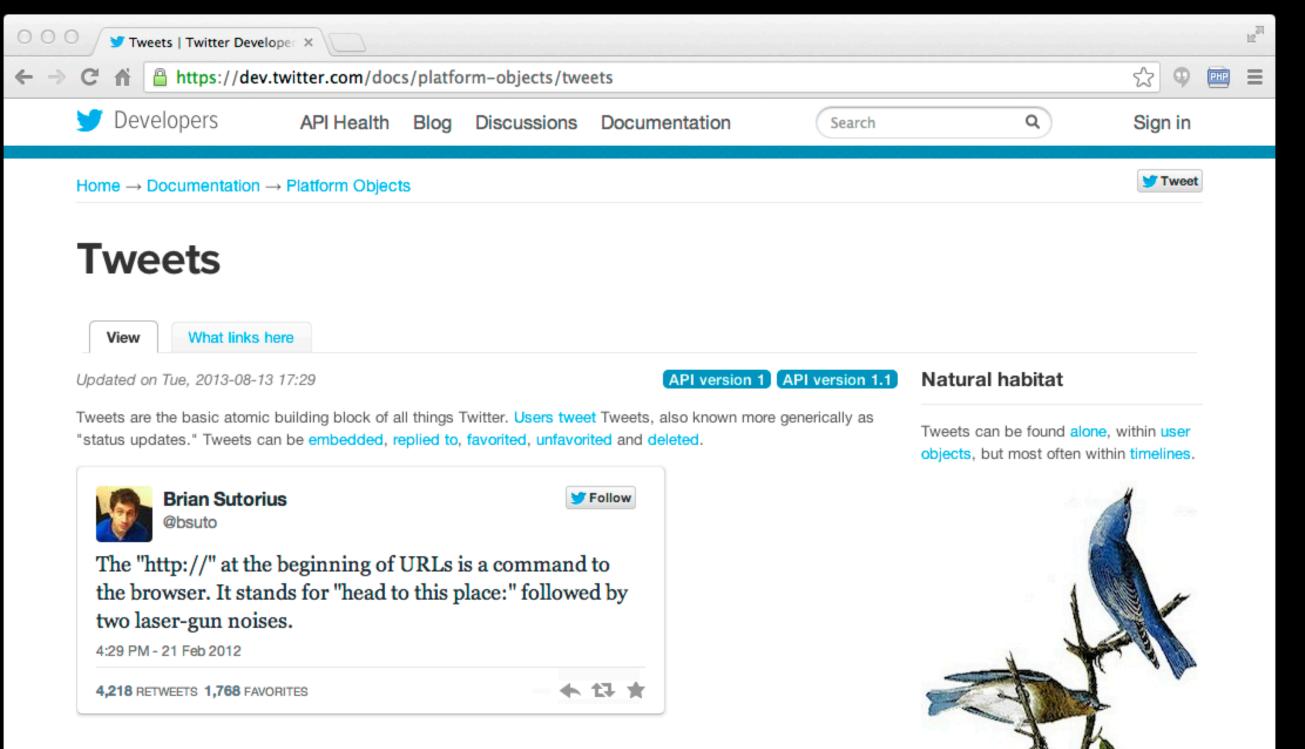
100,000 requests per day.

These limits are enforced to prevent abuse and/or repurposing of the Geocoding API, and may be changed in the future without notice. Additionally, we enforce a request rate limit to prevent abuse of the service. If you exceed the 24-hour limit or otherwise abuse the service, the Geocoding API may stop working for you temporarily. If you continue to exceed this limit, your access to the Geocoding API may be blocked.

The Geocoding API may only be used in conjunction with a Google map; geocoding results without displaying them on a map is prohibited. For complete details on allowed usage, consult the Maps API Terms of Service License Restrictions.



Moving from Basic Auth to OAuth →



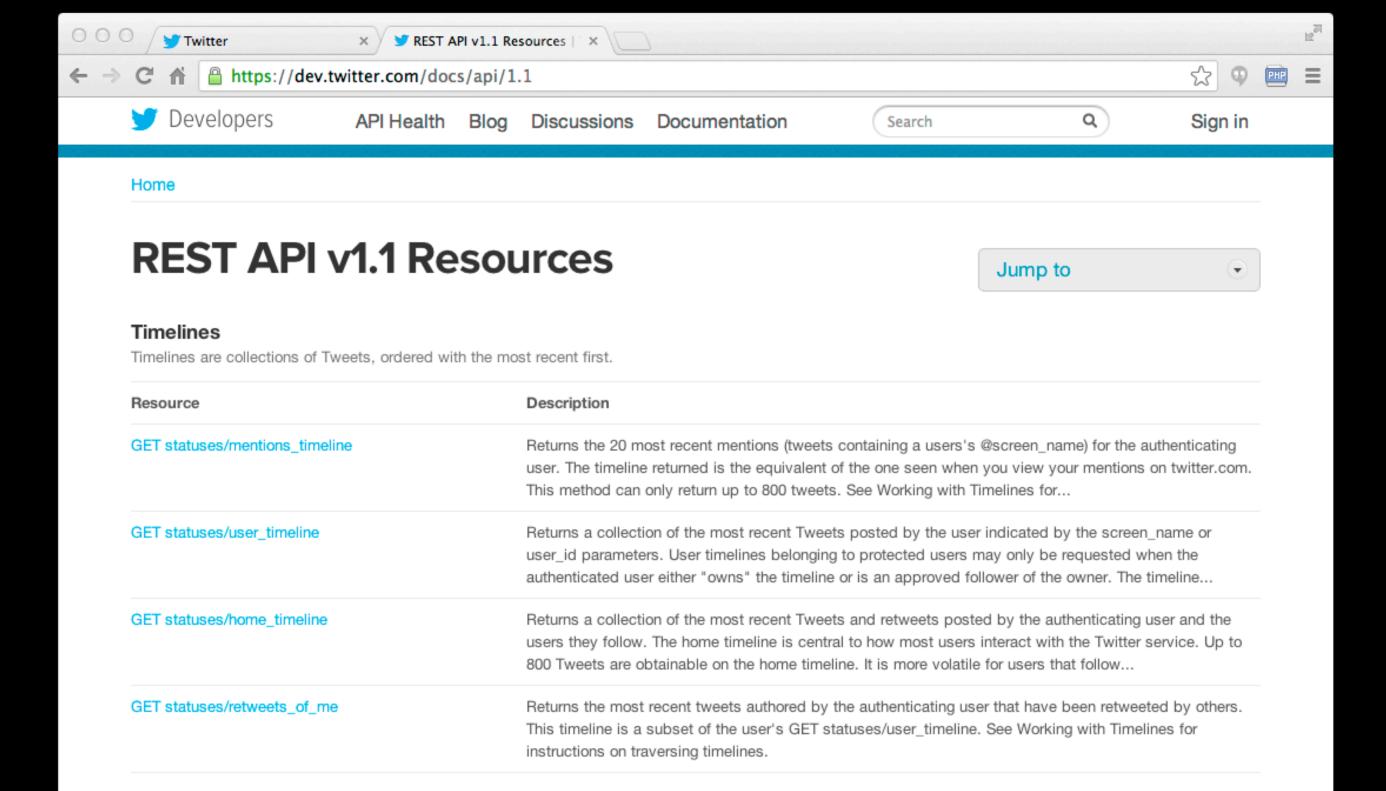
Field Guide

Consumers of Tweets should tolerate the addition of new fields and variance in ordering of fields with ease. Not all fields appear in all contexts. It is generally safe to consider a nulled field, an empty set, and the absence of a field as the same thing. Please note that Tweets found in Search results vary somewhat in structure from this document.

Field	Туре	Description
annotations	Object	Unused. Future/beta home for status annotations.

Related API Resources

GET favorites



Tweets

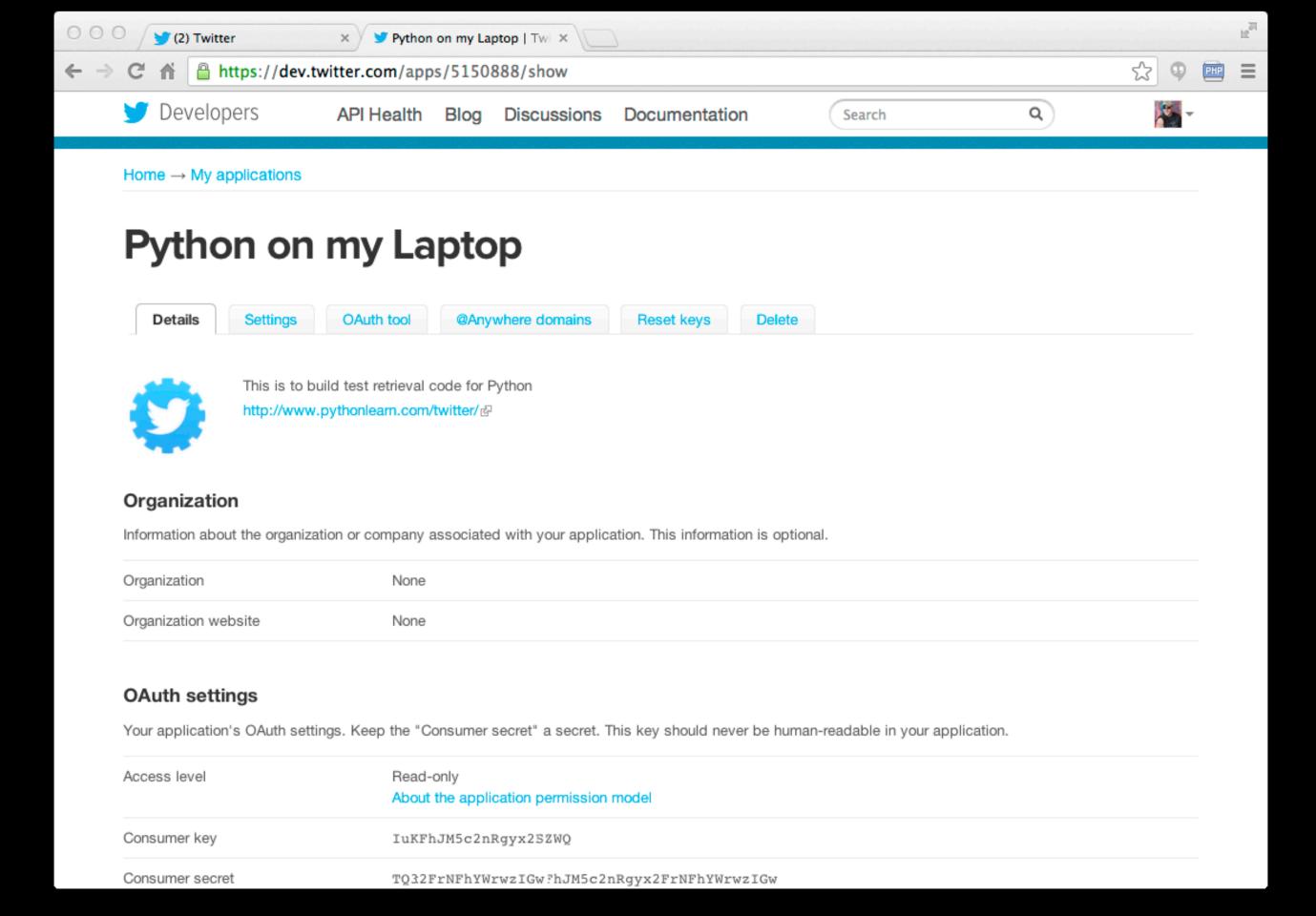
Tweets are the atomic building blocks of Twitter, 140-character status updates with additional associated metadata. People tweet for a variety of reasons about a multitude of topics.

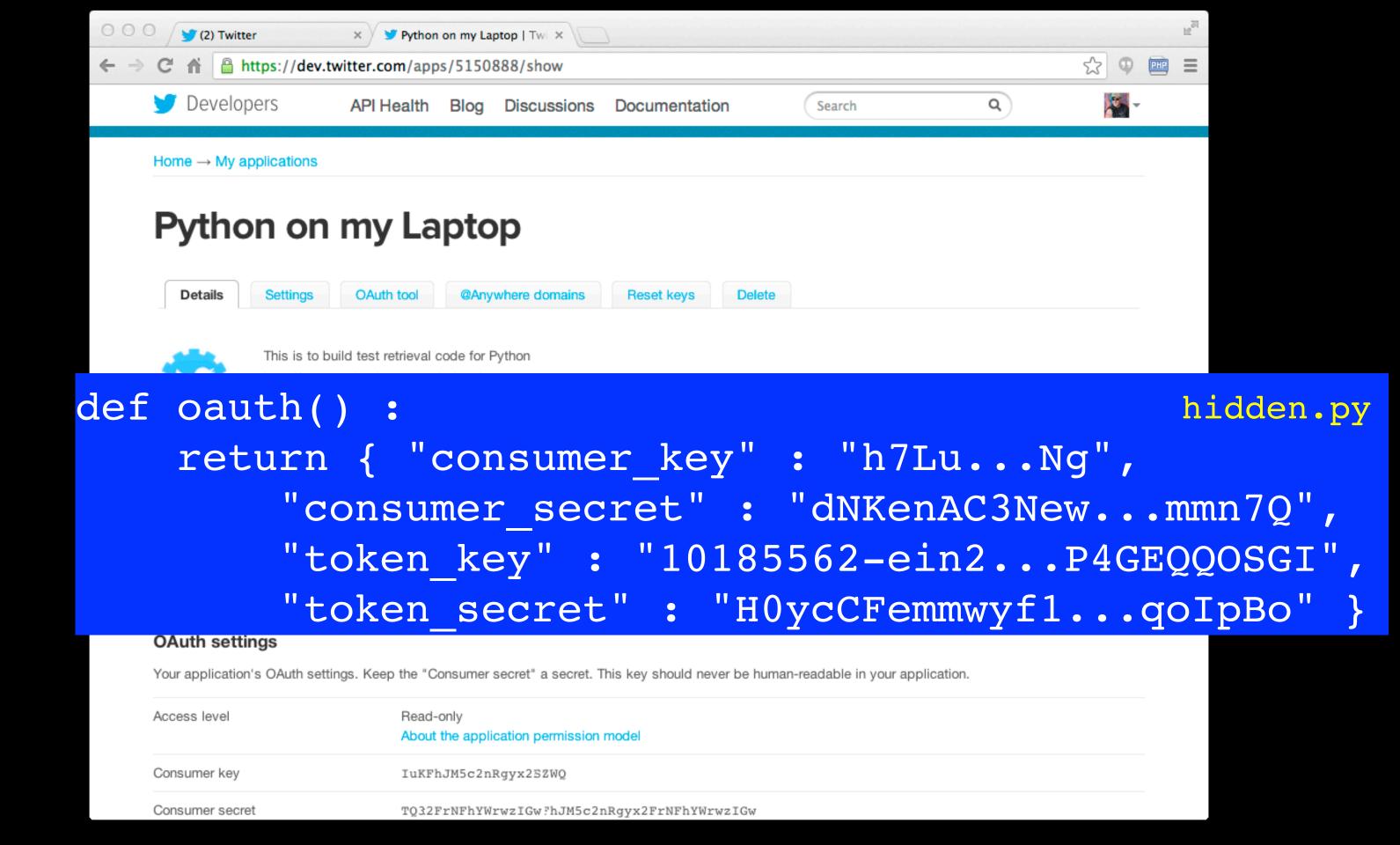
twitter2.py

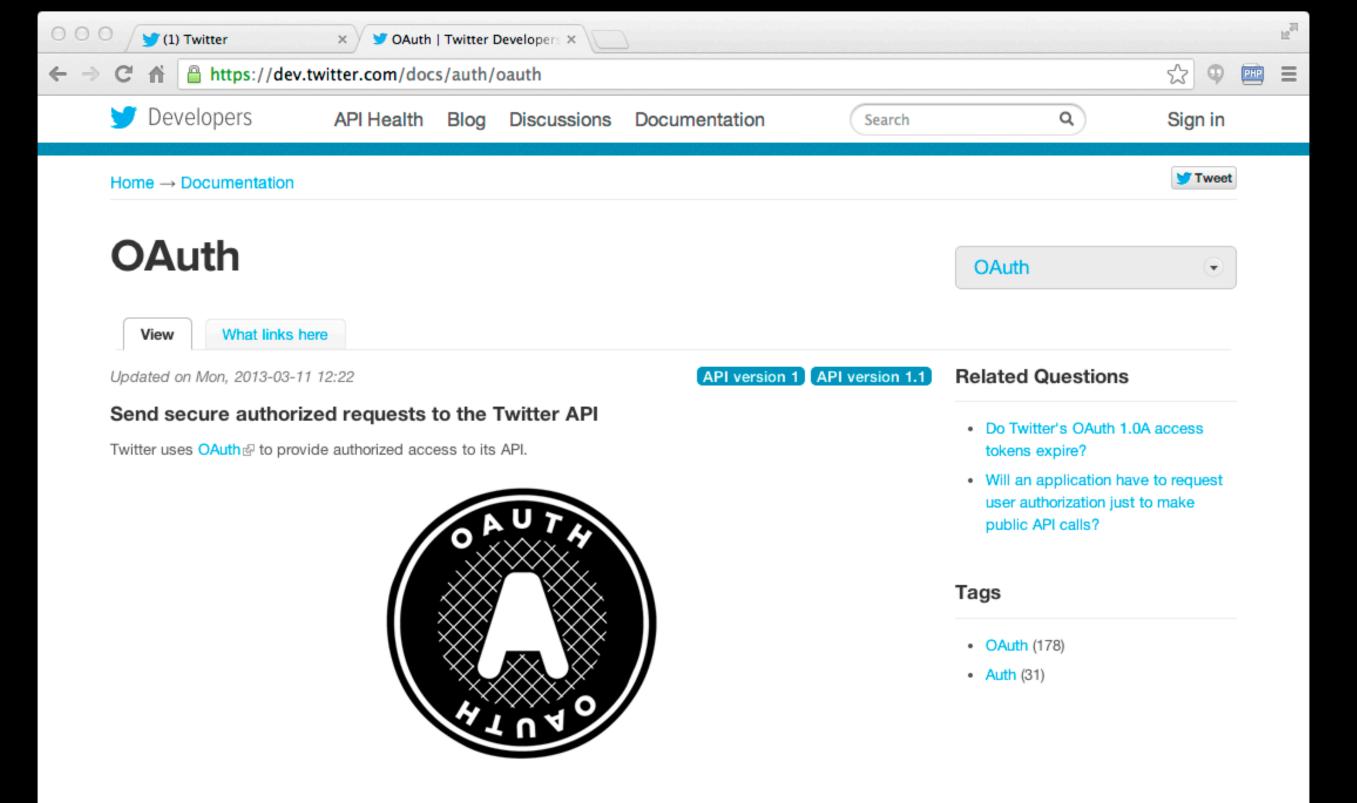
```
import urllib
import twurl
import json
TWITTER URL = 'https://api.twitter.com/1.1/friends/list.json'
while True:
    print ''
    acct = raw input('Enter Twitter Account:')
    if ( len(acct) < 1 ) : break</pre>
    url = twurl.augment(TWITTER URL,
        { 'screen name': acct, 'count': '5'} )
    print 'Retrieving', url
    connection = urllib.urlopen(url)
    data = connection.read()
    headers = connection.info().dict
    print 'Remaining', headers['x-rate-limit-remaining']
    js = json.loads(data)
    print json.dumps(js, indent=4)
    for u in js['users']:
        print u['screen name']
        s = u['status']['text']
        print ',s[:50]
```

```
Enter Twitter Account:drchuck
Retrieving https://api.twitter.com/1.1/friends ...
Remaining 14
    "users": [
            "status": {
                "text": "@jazzychad I just bought one .__.",
                 "created at": "Fri Sep 20 08:36:34 +0000 2013",
             },
             "location": "San Francisco, California",
             "screen name": "leahculver",
             "name": "Leah Culver",
         },
            "status": {
                "text": "RT @WSJ: Big employers like Google ...",
                 "created at": "Sat Sep 28 19:36:37 +0000 2013",
             },
             "location": "Victoria Canada",
             "screen name": " valeriei",
             "name": "Valerie Irvine",
     ],
Leahculver
   @jazzychad I just bought one .__._
Valeriei
   RT @WSJ: Big employers like Google, AT& T are h
Ericbollens
  RT @lukew: sneak peek: my LONG take on the good &a
halherzog
 Learning Objects is 10. We had a cake with the LO,
```

twitter2.py







Features

- . Secure Users are not required to share their passwords with 3rd party applications, increasing account security.
- Standard A wealth of client libraries and example code are compatible with Twitter's OAuth implementation.

Direct dia Arctionationation Mandal

https://api.twitter.com/1.1/statuses/user timeline.json?count=2&oau

Summary

- Service Oriented Architecture allows an application to be broken into parts and distributed across a network
- An Application Program Interface (API) is a contract for interaction
- Web Services provide infrastructure for applications cooperating (an API) over a network - SOAP and REST are two styles of web services
- XML and JSON are serialization formats