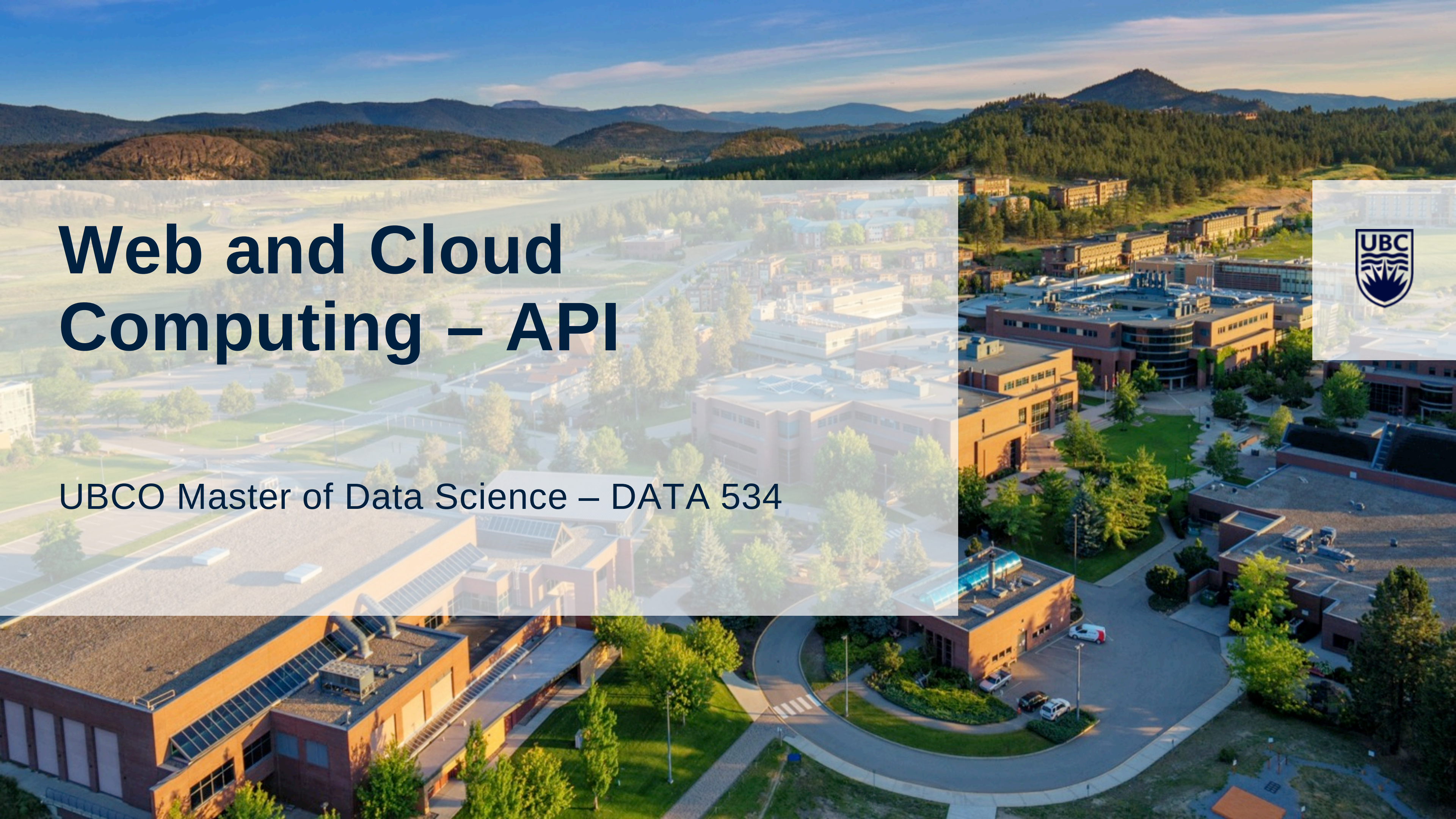


Web and Cloud Computing – API

UBCO Master of Data Science – DATA 534

Lecture Learning Goals

- Describe key benefits and limitations of web scraping.
- Be able to describe what APIs are and how they are used.
- Be able to use a common web API (e.g., yahoo, twitter, etc.) to collect data and analyze it in an interesting way.
- Describe the benefits and limitations of API’s compared to web scraping.

Limitations of Scraping

Is scraping the best way to get data?

Limitations

- Dynamic Websites
- Extracting huge amounts of data
- Extracting data from non-HTML content.
- changes of the page layout.
- Dealing with Infinitive Scrolling/Load More

Image from dribble.com by artrayd

Scraping Dynamic websites

- BeautifulSoup or Requests don't automatically fetch dynamic content from a web page

Other Options:

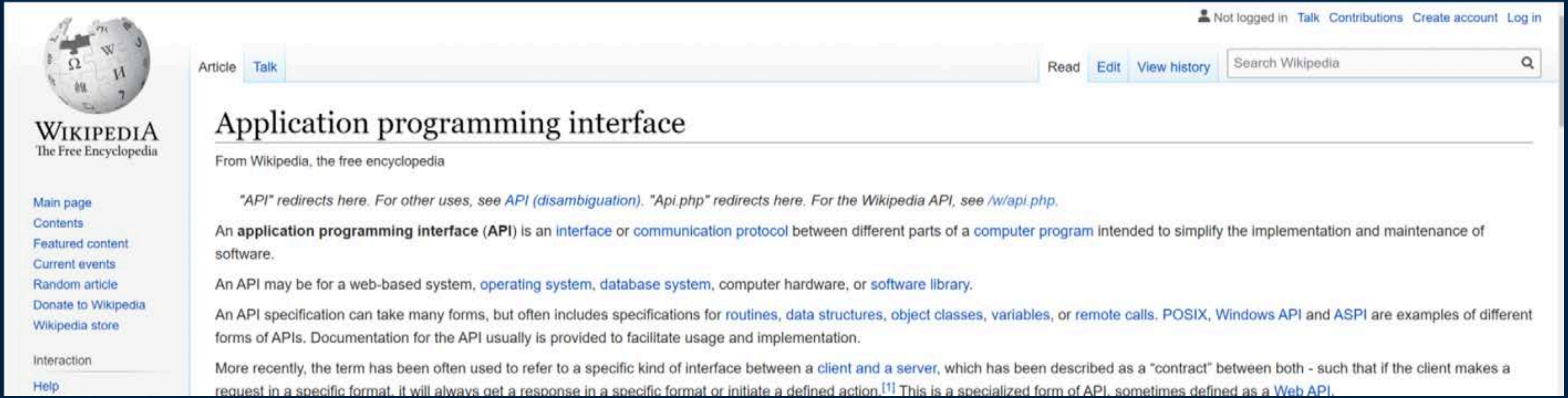
- Selenium
- Ajax requests

API

Application

Programming

Interface



Application programming interface

From Wikipedia, the free encyclopedia

"API" redirects here. For other uses, see API (disambiguation). "Api.php" redirects here. For the Wikipedia API, see /w/api.php.

An **application programming interface** (**API**) is an interface or communication protocol between different parts of a computer program intended to simplify the implementation and maintenance of software.

An API may be for a web-based system, operating system, database system, computer hardware, or software library.

An API specification can take many forms, but often includes specifications for routines, data structures, object classes, variables, or remote calls. POSIX, Windows API and ASPI are examples of different forms of APIs. Documentation for the API usually is provided to facilitate usage and implementation.

More recently, the term has been often used to refer to a specific kind of interface between a client and a server, which has been described as a "contract" between both - such that if the client makes a request in a specific format, it will always get a response in a specific format or initiate a defined action.[1] This is a specialized form of API, sometimes defined as a Web API.

API

It’s a software intermediary that allows two applications to talk to each other.

Every time you use the weather app in your phone, send a mobile payment, or turn on the lights at home from your phone, you’re using an API.

APIs allow different application to complement each other. They allow you to save time and help not to invent the wheel again.

Types of API

- Open APIs
  - no restrictions to access
- Partner APIs
  - Specific rights or licenses to access
- Internal APIs
  - Just internal systems can use it
- Composite APIs
  - Combines different data and service API

API Protocols and Architecture

1) XML-RPC

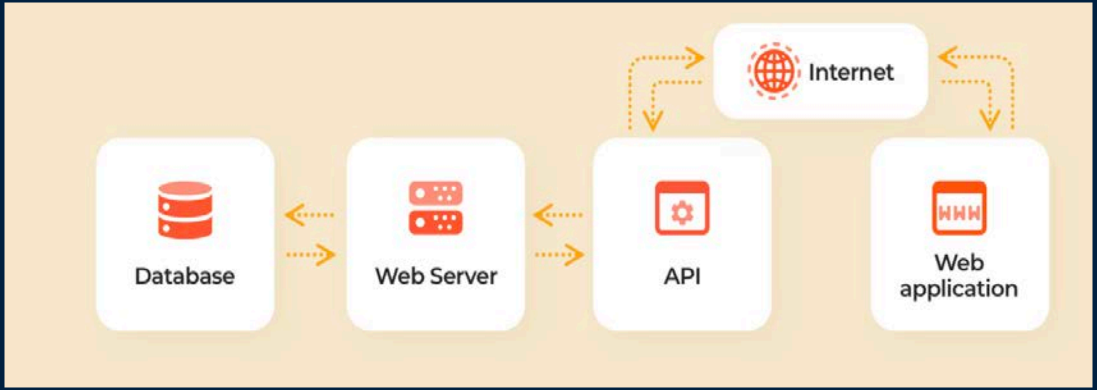
2) JSON-RPC

3) SOAP

- (simple Object Access Protocol)

4) REST

- Representation State Transfer



Source: https://www.cleveroad.com/images/article-previews/40ca78a7a9db7adfb6bb861fc6b8910ae2ef4bb79f5508007d166f01df5c1038.png

REST(ful) Representation State Transfer

REST API follows a set of rules to communicate with an app with a service.

REST API is referred to as RESTful API

Characteristics:

- client-server, typically HTTP based, stateless server, flexible

Characteristics of a REST based network

- Client-Server: a pull-based interaction style (Client request data from servers as and when needed).

- Stateless: each request from client to server must contain all the information necessary to understand the request, and cannot take advantage of any stored context on the server.

- Cache: to improve network efficiency, responses must be capable of being labeled as cacheable or non-cacheable.

Characteristics of a REST based network

- Uniform interface: all resources are accessed with a generic interface (e.g., HTTP GET, POST, PUT, DELETE).

- Named resources -the system is comprised of resources which are named using a URL.

- Interconnected resource representations - the representations of the resources are interconnected using URLs, thereby enabling a client to progress from one state to another.

Rest –An architectural Style

Elements

- Components – Proxy , gateway etc
- Connectors – client , server etc
- Data – resource , representation etc

Resource

- Anything that can be named and lives on a server.
- static file, a document, a database, a picture…
- The resource identifier is the specific location of the resource requested

Rest –An architectural Style

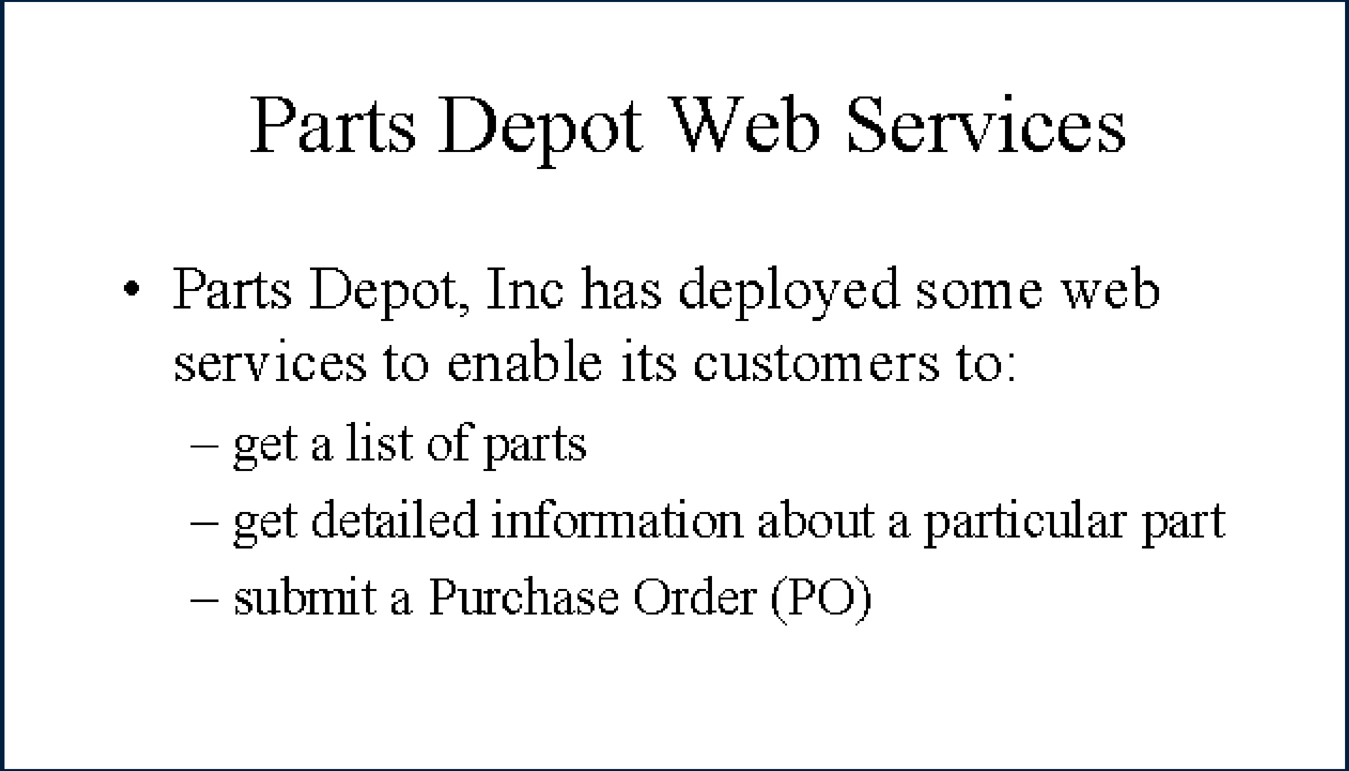
Elements

- Components – Proxy , gateway, origin server, etc
- Connectors – client , server, cache, etc
- Data – resource , representation etc

Representation

- Data sent to the client.
- The data is not processed on the server, but rather interpreted by the client.
- XML and JSON are the most popular resource representations.

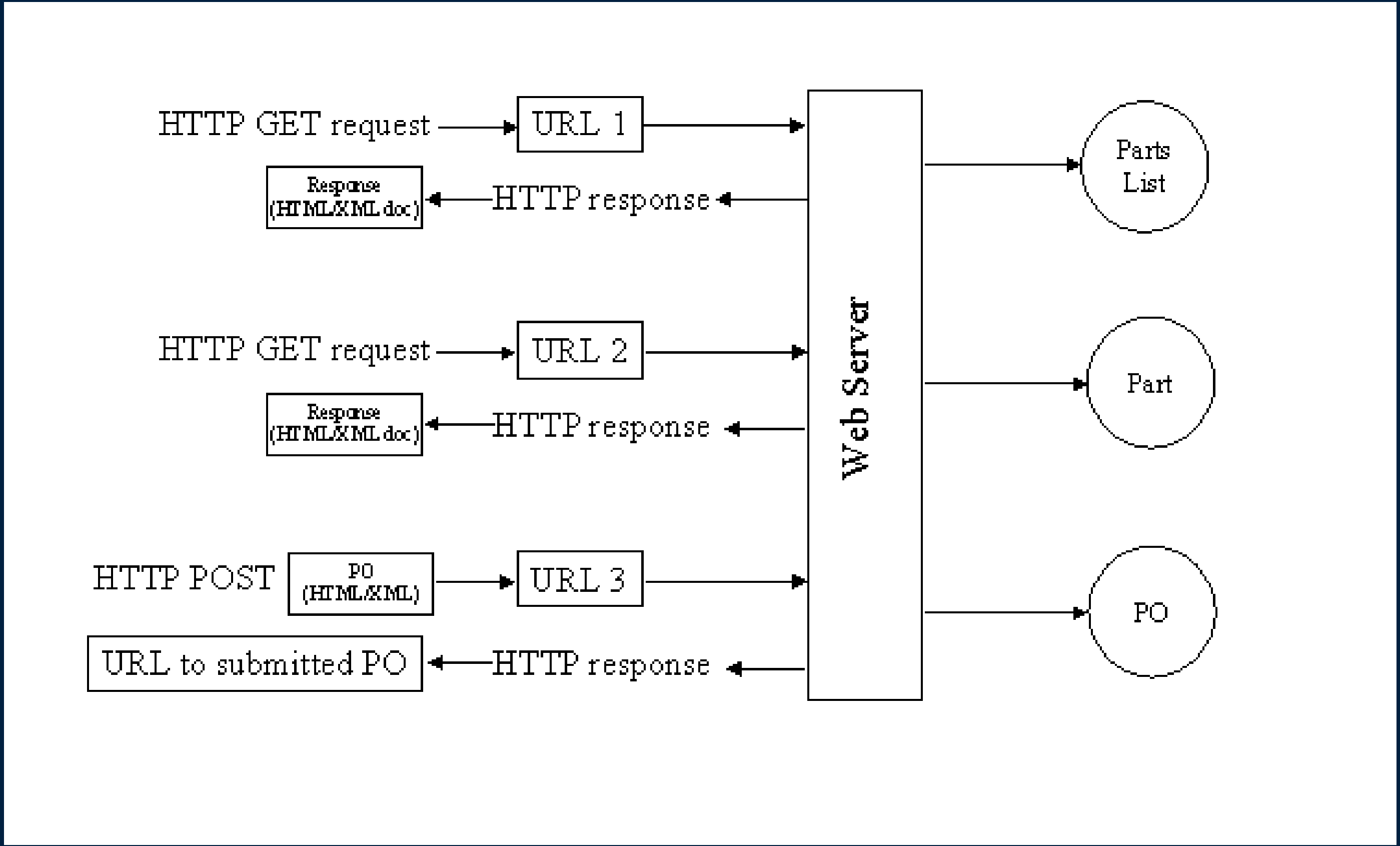
EXAMPLE:



Parts Depot Web Services

- Parts Depot, Inc has deployed some web services to enable its customers to:
  - get a list of parts
  - get detailed information about a particular part
  - submit a Purchase Order (PO)

REST way of Implementing the web services



Service –Get parts list

The web service makes available a URL to a parts list resource

Client uses : http://www.parts-depot.com/parts

Document Client receives :

```
<?xml version="1.0"?>
<p:Parts xmlns:p=http://www.parts-depot.com xmlns:xlink="http://www.w3.org/1999/xlink">
        <Part id="00345" xlink:href="http://www.parts-depot.com/parts/00345"/>
        <Part id="00346" xlink:href="http://www.parts-depot.com/parts/00346"/>
        <Part id="00347" xlink:href="http://www.parts-depot.com/parts/00347"/>
        <Part id="00348" xlink:href="http://www.parts-depot.com/parts/00348"/>
</p:Parts>
```

Service –Get detailed part data

The web service makes available a URL to each part resource.

- Client uses: http://www.parts-depot.com/parts/00345

Document Client receives :

```
<?xml version="1.0"?>
<p:Part xmlns:p="http://www.parts-depot.com" xmlns:xlink="http://www.w3.org/1999/xlink">
		<Part-ID>00345</Part-ID>
		<Name>Widget-A</Name>
		<Description>This part is used within the frap assembly</Description>
		<Specification xlink:href="http://www.parts-depot.com/parts/00345/specification"/>
<UnitCostcurrency="USD">0.10</UnitCost>
		<Quantity>10</Quantity>
</p:Part>
```

Web APIs

Some web site’s provide direct access to their data

Why would they do this?

Why would some web sites not do this?

Some Examples

Twitter (https://developer.twitter.com/en/docs/twitter-api)

Translink (https://www.translink.ca/about-us/doing-business-with-translink/app-developer-resources)

Github (https://docs.github.com/en/rest?apiVersion=2022-11-28)

.

.

.

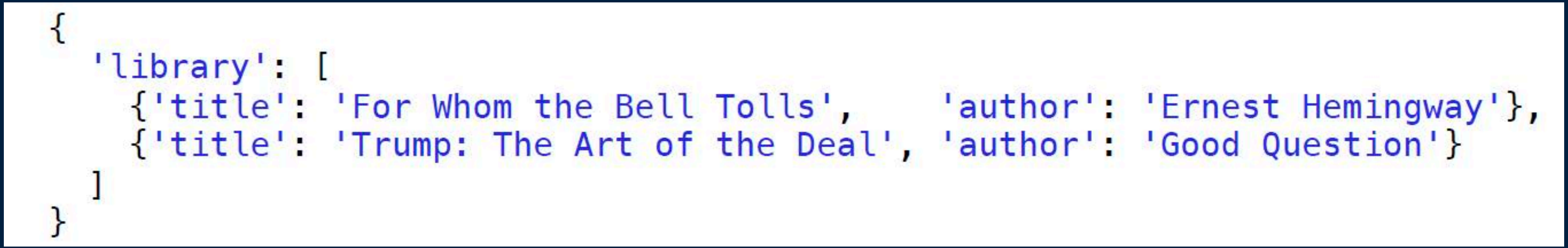
How do you get started ?

- API Documentation
- Registration and authentication
- Examples through your browser

Transiting Data

JSON

- textual description of python (javascript actually) objects
- arrays and dictionaries

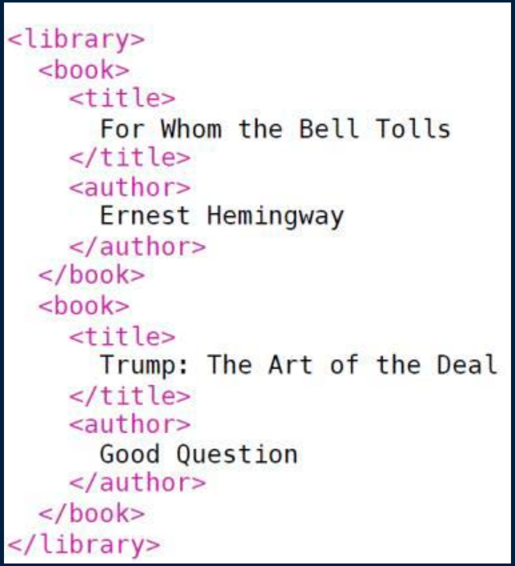


```
{
  'library': [
    {'title': 'For Whom the Bell Tolls',    'author': 'Ernest Hemingway'},
    {'title': 'Trump: The Art of the Deal', 'author': 'Good Question'}
  ]
}
```

Transiting Data

XML

- hierarchical description of tagged data



```
<library>
  <book>
    <title>
      For Whom the Bell Tolls
    </title>
    <author>
      Ernest Hemingway
    </author>
  </book>
  <book>
    <title>
      Trump: The Art of the Deal
    </title>
    <author>
      Good Question
    </author>
  </book>
</library>
```

Using a Web API

Provider defines

message format for requests and responses

- usually in both XML and JSON

registration and authentication

- usually using Oauth

Language integration

- might be provided or you might have to do it yourself
- if provided, usually someone other than data source
- library API for various languages like python
- you write a python program that calls library procedures
- library formats messages, sends them to web provider, translates responses as return values 10

Getting JSON Data

Need to select output format using API

- e.g., http header: accept = application/json

View in browser or Postman

- good for exploration / debugging

Use request .get

- this returns a python array or dictionary

Get a string and parse

- import json
- x = json.loads(aJSONString)

Examples

Cocktail DB

```
import requests
from bs4 import BeautifulSoup
import json

url = "https://www.thecocktaildb.com/api/json/v1/1/lookup.php?iid=552"
r = requests.get(url)
s = BeautifulSoup(r.text,'html5lib')
site_json=json.loads(s.text)
desc = site_json['ingredients'][0]['strDescription']
print(desc)
```

OUPUT:

```
Elderflower cordial is a soft drink made largely from a refined sugar and water solution and uses the flowers of the European
elderberry. Historically the cordial has been popular in North Western Europe where it has a strong Victorian heritage.
```

Cocktail DB

Search cocktail by name
https://www.thecocktaildb.com/api/json/v1/1/search.php?s=margarita

List all cocktails by first letter
https://www.thecocktaildb.com/api/json/v1/1/search.php?f=a

Search ingredient by name
https://www.thecocktaildb.com/api/json/v1/1/search.php?i=vodka

Lookup full cocktail details by id
https://www.thecocktaildb.com/api/json/v1/1/lookup.php?i=11007

Lookup ingredient by ID
https://www.thecocktaildb.com/api/json/v1/1/lookup.php?iid=552

Lookup a random cocktail
https://www.thecocktaildb.com/api/json/v1/1/random.php

Translink

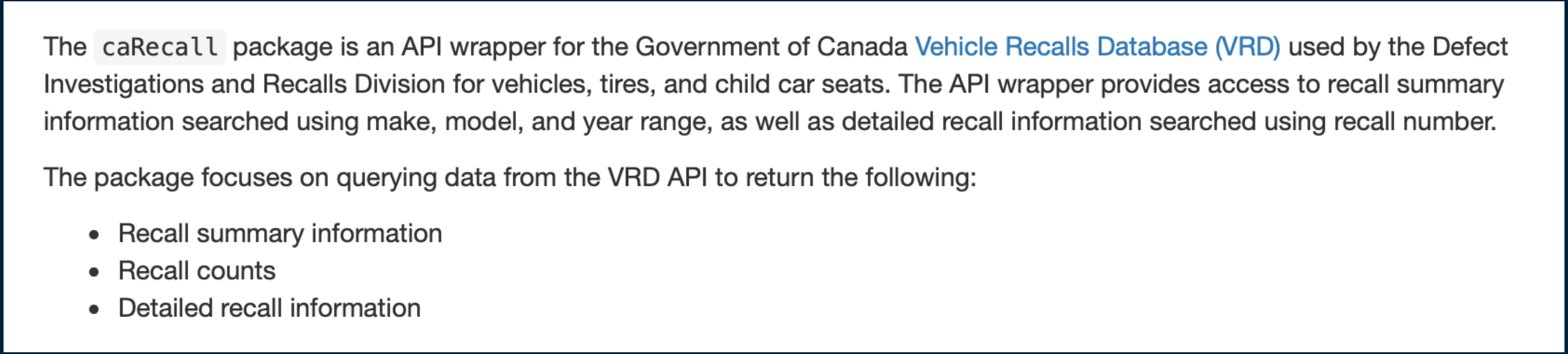
```
import requests


# get your own api token from developer.translink.ca
apikey = '?';


x = 
requests.get('http://api.translink.ca/rttiapi/v1/stops/61935?apikey={}'
.format(apikey),headers={'accept': 'application/JSON'}).json()
```

Next Lecture

Internet and Web Infrastructure

THE UNIVERSITY OF BRITISH COLUMBIA

Example: caRecall



The `caRecall` package is an API wrapper for the Government of Canada Vehicle Recalls Database (VRD) used by the Defect Investigations and Recalls Division for vehicles, tires, and child car seats. The API wrapper provides access to recall summary information searched using make, model, and year range, as well as detailed recall information searched using recall number.

The package focuses on querying data from the VRD API to return the following:

- Recall summary information
- Recall counts
- Detailed recall information

https://wraysmith.github.io/caRecall/