Plan for Today

- Reading Quiz
- APIs
- JSON
- MySQL + JSON
Readings for Today

- Chapter 13 from our Data Wrangling text
- JSON Data Interchange Format
Question 1

What is an API?

A. It’s a service for pulling data
B. It’s a service for pushing data
C. It’s a service for searching data
D. It’s a service that outputs responses in JSON format
E. All of the above
Question 2

A REST API client uses HTTP GET to pull data from the API service and HTTP POST to push data to the API service:

A. True

B. False
Question 3

Which of the following statements is false:

A. JSON is a collection of name/value pairs
B. JSON is a self-describing data format
C. JSON is used mostly with JavaScript
D. JSON is used for data exchange
E. JSON can represent nested data
Question 4

Which of the following is a **valid** JSON construct:

A. Array
B. Loop
C. Table
D. Function
E. Parameter
Question 5

How many nested objects are in this JSON document?

```json
{
    "text": "RT @hiangieee: The students of @UTAustin took it to the streets today. #utprotest #Morningafter https://t.co/jvTLajTqT8",
    "retweet_count": 6236,
    "id_str": "796541058192736256",
    "favorited": false,
    "user": {
        "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/569242807726915584/WDmWx2F.jpeg",
        "entities": {
            "description": {
                "urls": []
            }
        },
        "followers_count": 73,
        "utc_offset": -21600
    },
    "metadata": {
        "iso_language_code": "en",
        "result_type": "recent"
    }
}
```
JSON Format

```json
{
  "firstName": "Jason",
  "middleInitial": "M",
  "lastName": "Jones",
  "address": {
    "street": "444 Guadalupe",
    "city": "Austin",
    "state": "TX",
    "zip": 78705
  },
  "email": [
    "jason@utexas.com",
    "jjones@gmail.com"
  ],
  "phone": [
    {"work": "512-555-1212"},
    {"cell": "512-222-1234"}
  ]
}

Source: http://www.json.org
```
Concept Question 1

Suppose we wanted to add a second person to this document. How can we modify the JSON structure to represent $n$ people?

```json
{
    "firstName": "Jason",
    "middleInitial": "M",
    "lastName": "Jones",
    "address": {
        "street": "444 Guadalupe",
        "city": "Austin",
        "state": "TX",
        "zip": 78705
    },
    "email": [
        "jason@utexas.com",
        "jjones@gmail.com"
    ],
    "phone": [
        {"work": "512-555-1212"},
        {"cell": "512-222-1234"}
    ]
}
```

A. Add a Person array, each person = an element in the array and each element = a nested object

B. Make each person = a nested object. The individual person objects are separated by a ", , "

C. B + add a label for each person object. The label = a unique identifier such as an ssn.

D. All of the above

E. None of the above
Anatomy of a (partial) Tweet

Twitter API Field Guide: https://dev.twitter.com/overview/api/tweets

JSON Editor Online: https://chrome.google.com/webstore to download the Chrome plug-in.
UT Class Enrollment & Twitter

Logical ERD - UT Class Enrollment - CS 327E Fall 2016
New DDL

```sql
drop table if exists Major;
create table Major (
    code int auto_increment primary key,
    name varchar(32) not null,
    college varchar(32) not null
);

alter table Student add column major_code int;
alter table Student add constraint fk_major_code
    foreign key (major_code) references Major(code);

drop table if exists Tweet;
create table Tweet (  
    tweet_id varchar(32) generated always  
        as (json_unquote(json_extract(tweet_doc, '\$.id_str'))) stored primary key,  
    screen_name varchar(32) generated always  
        as (json_unquote(json_extract(tweet_doc, '\$.user.screen_name'))) stored,  
    created_at datetime generated always  
        as (str_to_date(json_unquote(json_extract(tweet_doc, '\$.created_at')),  
            '%a %b %d %H:%i:%s +0000 %Y')) stored,  
    tweet_doc json,  
    major_code int,  
    foreign key (major_code) references Major(code)
);
```
Concept Question 2

Suppose we want to store the `favorite_count` from a tweet in its own field because we plan to query it frequently. How can we formulate the JSON path expression to extract the `favorite_count` value from a tweet?

Path expressions:

A. $.contributors.favorite_count
B. $.favorite_count
C. $.text.favorite_count
D. $.id.favorite_count
tweepy = Twitter API wrapper, makes OAuth a lot simpler

Visit https://apps.twitter.com to obtain API key and secret along with TOKEN key and secret for your application. Note: API key is also called Consumer key.
What does this code sample do?

```
from tweepy import Cursor
from tweepy import OAuthHandler

auth = tweepy.OAuthHandler(API_KEY, API_SECRET)
auth.set_access_token(TOKEN_KEY, TOKEN_SECRET)

api = tweepy.API(auth)
query = '#childlabor'
cursor = tweepy.Cursor(api.search, q=query, lang="en")
for page in cursor.pages():
    for item in page:
        store_tweet(item)
```

A. Searches the database for tweets related to “child labor”.

B. Searches Twitter for tweets containing #childlabor, pulls those tweets, and stores each one into its own table in SQLite.

C. Answer B except that the tweets are all stored in the same tweets table.

D. Searches Twitter for tweets containing #childlabor and generates a web page for each tweet.

E. None of the above.
```python
def do_data_pull(api_inst):
    sql_query = "select code, name from Major order by name"

    try:
        conn = create_connection()
        db_cursor = conn.cursor()
        query_status = run_stmt(db_cursor, sql_query)
        resultset = db_cursor.fetchall()

        for record in resultset:
            major_code = record[0]
            major_name = record[1]

            utexas_query = "(#UTexas OR @UTAustin OR url:utexas.edu) AND "
            twitter_query = utexas_query + "," + major_name + ""

            print "twitter_query: " + twitter_query
            twitter_cursor = tweepy.Cursor(api_inst.search, q=twitter_query, lang="en")

            for page in twitter_cursor.pages():
                for item in page:
                    json_str = json.dumps(item._json)
                    print "found a " + major_name + " tweet"
                    insert_stmt = "insert into Tweet(tweet_doc, major_code) values(\%s, \%s)"
                    run_prepared_stmt(db_cursor, insert_stmt, (json_str, major_code))
                    do_commit(conn)

    except pymysql.Error as error:
        is_success = False
        print "do_data_pull: " + e.strerror
```

Plan for Next Week

- Last Quiz on Monday: Readings will come straight from the MySQL Reference Guide. See class web page for details.
- Final Project: Assignment and rubric will be out on Monday.
- No class on Wednesday: Thanksgiving break!