



上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

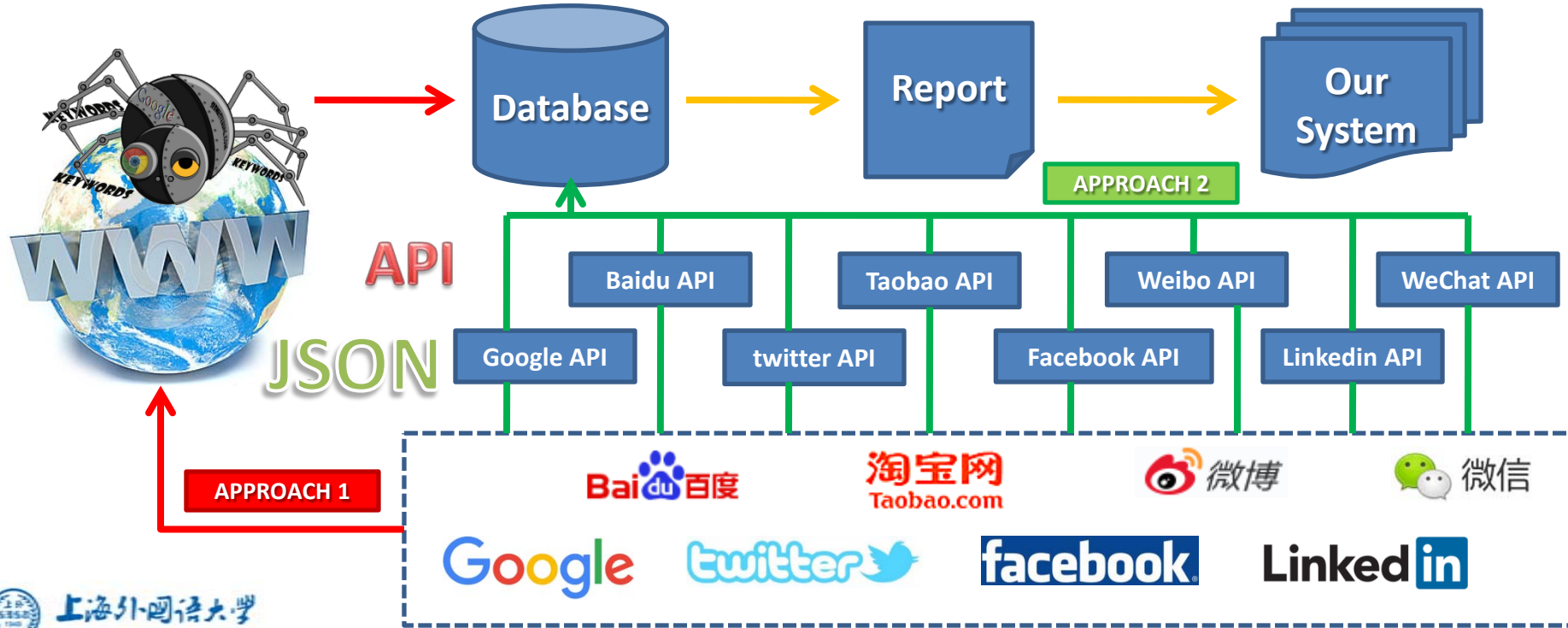
New Media Data Analytics and Application

Lecture 5: Data Collection via Web API

Ting Wang

Objectives

How to get data from social media?



- Web API
- JSON
- Examples of Web API Applications
 - WeChat
 - Film Box Office by Juhe
 - douban





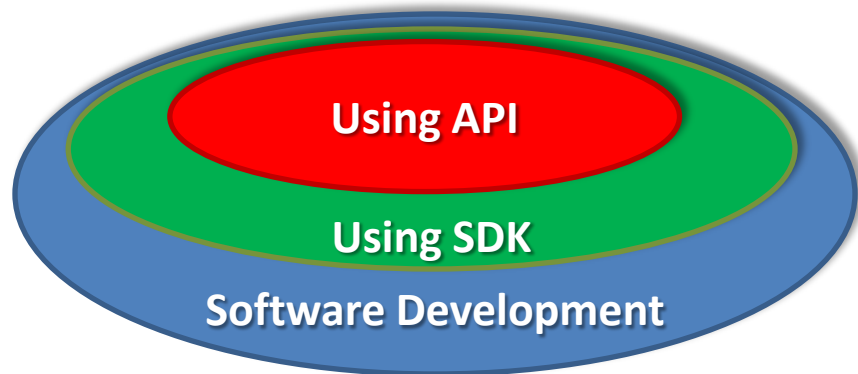
an access to the web site data

Web API

SDK (Software Development Kit)

软件开发工具包

a set of
software development tools,
source code,
instructions,
examples,
and ...



How to print a booklet for our sport games?



2

Using SDK
在人家的套路基础上走

1



蔡伦



毕昇

Traditional Software
Development
自己设计套路走

3

Using Web API
自己不动脑子，
完全按人家的套路走

淘宝网 Taobao.com 宝贝 彩页设计印刷 搜索 在结果中排除 请输入要排除的词 确定

| 商家 | 促销/服务 | 价格/规格 |
|------|----------------------|---|
| 益高印刷 | 免设计费 三皇冠厂家 全国包邮 | 500张 25元 1000张 30元 2000张 80元 3000张 90元 5000张 150元 |
| 益好 | 足克足量 定稿后次日发货 | 500张 45元 1000张 70元 2000张 148元 3000张 218元 5000张 298元 |
| 东讯 | 五折促销 A4宣传单/海报印刷 免费设计 | 500张 45元 1000张 70元 2000张 148元 3000张 218元 5000张 298元 |
| 源拓印刷 | A4宣传单印刷 免费设计 足克足量 | 500张 40元 1000张 70元 2000张 148元 |

画册印刷 买贵退差价

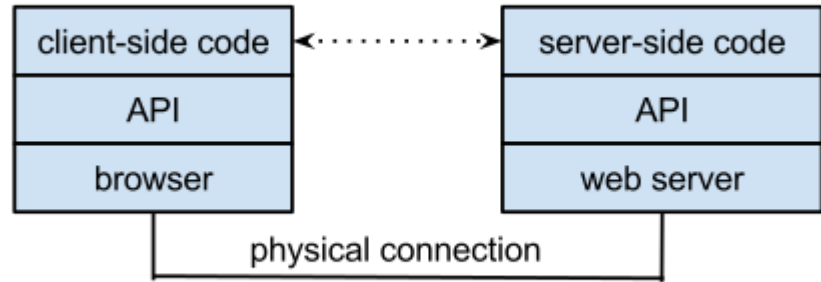
| 商家 | 价格/规格 |
|------|-----------------|
| 益高印刷 | ¥68.00起 1350人付款 |
| 益好 | ¥68.00起 7583人付款 |
| 东讯 | ¥68.00起 5745人付款 |
| 源拓印刷 | ¥68.00起 5240人付款 |

API (Application Programming Interface)

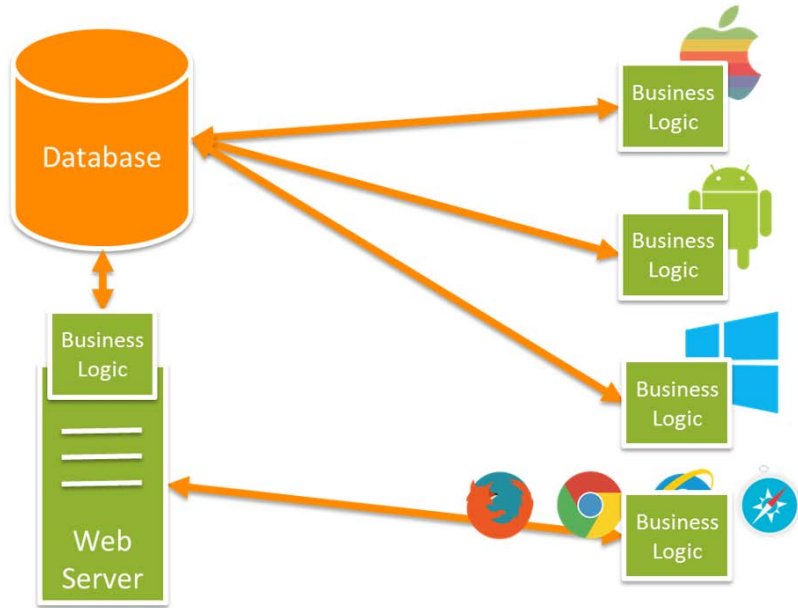
应用程序接口

a set of
subroutine definitions,
protocols,
and tools

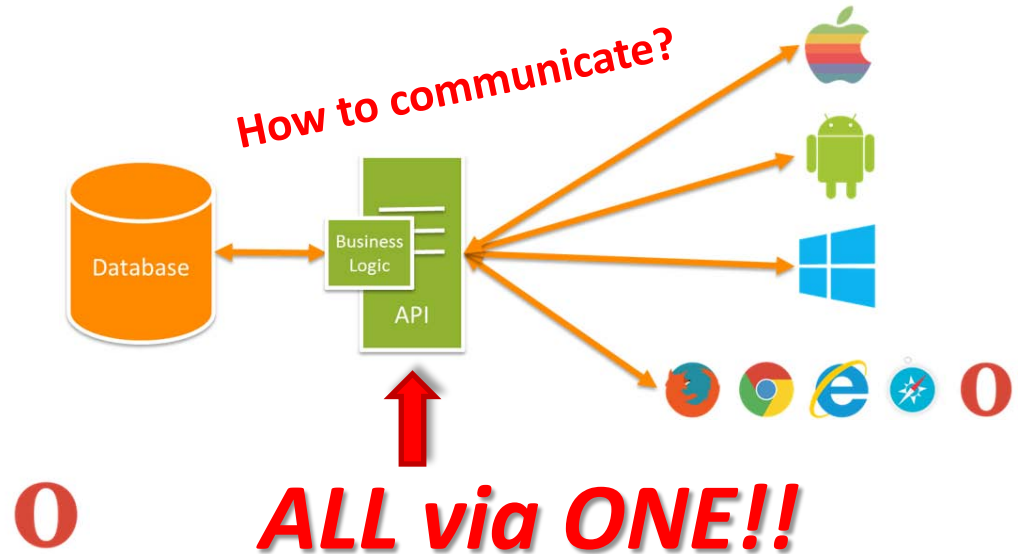
- What is Web API?



Short Short Ago



Now



Integrating all the operations into one package, including “insert”, “delete”, “update”, and “select”.

In Unified Style, two crucial elements:

– Unified Method :

- HTTP (HyperText Transfer Protocol)

超文本传输协议

– Unified Language :

- JSON
- XML



HTTP(HyperText Transfer Protocol)

超文本传输协议

an application protocol for distributed, collaborative, hypermedia information systems, which is the foundation of data communication for the World Wide Web.

- GET
 - Give Me Information
- POST
 - Give Server Information

Do you remember
“Login.py” ?



Web API

```
1 #!/usr/bin/env python3
2 # -*- coding: utf-8 -*-
3
4 from flask import Flask
5 from flask import request
6 import pymysql
7 import pymysql
8 import sqlite3
9
10 app = Flask(__name__)
11
12 @app.route('/', methods=['GET', 'POST'])
13 def home():
14     return '''<h1>Home</h1>
15     <p><a href="/login">Login</a></p>'''
16
17
18 @app.route('/login', methods=['GET'])
19 def login_form():
20     return '''<form action="/login" method="POST">
21     <p><input name="username"></p>
22     <p><input name="password" type="password"></p>
23     <p><button type="submit">Log In Now</button></p>
24     </form>'''
25
26 @app.route('/login', methods=['POST'])
27 def login():
28     # get data from database:
29     # 1.Connection Open
30     conn = pymysql.connect(user='sa', password='123456', database='login')
31     # 2.Cursor Creating:
32     cursor = conn.cursor()
33     # 3.SQL Execution
34     sqlstr = 'select * from customer where customer_name=\'+request.form['username']+\'
35     |
36     | \'+request.form['password']+\'
37     # 4.Cursor Moving
38     cursor.execute(sqlstr)
39     numrows = len(cursor.fetchall())
40     conn.commit()
41     # 5.Connection Close
42     cursor.close()
43     conn.close()
44     # get data from request:
45     if numrows>0:
46         return '<h3>Hello, '+request.form['username']+\'</h3>'
47     return '<h3>Bad username or password.</h3>'
48
49 if __name__ == '__main__':
50     app.run()
```



Language of Web API

– XML (EXtensible Markup Language)

```
<?xml version="1.0" encoding="UTF-8" ?>
  <employees>
    <firstName>Bill</firstName>
    <lastName>Gates</lastName>
  </employees>
  <employees>
    <firstName>George</firstName>
    <lastName>Bush</lastName>
  </employees>
  <employees>
    <firstName>Thomas</firstName>
    <lastName>Carter</lastName>
  </employees>
```



Ref. <http://www.w3school.com.cn/x.asp>



Language of Web API

– JSON (JavaScript Object Notation)

```
{ "employees": [  
  { "firstName":"Bill" , "lastName":"Gates" },  
  { "firstName":"George" , "lastName":"Bush" },  
  { "firstName":"Thomas" , "lastName":"Carter" }  
]}
```

Ref. <http://www.w3school.com.cn/json/>

Transfer between XML and JSON

<http://www.bejson.com/xml2json/>

在线XML、JSON数据互转
如题

| XML | JSON |
|--|---|
| <pre><?xml version="1.0" encoding="UTF-8" ?> <employees> <firstName>Bill</firstName> <lastName>Gates</lastName> </employees> <employees> <firstName>George</firstName> <lastName>Bush</lastName> </employees> <employees> <firstName>Thomas</firstName> <lastName>Carter</lastName> </employees></pre> | <pre>{ "employees": [{ "firstName": "Bill", "lastName": "Gates" }, { "firstName": "George", "lastName": "Bush" }, { "firstName": "Thomas", "lastName": "Carter" }] }</pre> |

转义编码

→

←

清空





Ask A Question

Why JSON is more and more popular?



EXAMPLE 1:
Weather Forecast 1

I ♥ APIs

Get this free sticker at apigee.com

Tips:

1. Register, Login, and get your **API KEY** from each website before you use API.
2. Different websites have different API Keys.

Step 1: Register and get your API Key



1

A registration form for an API key. It has two radio buttons for '类型' (Type): '个人' (Personal) and '公司' (Company). Below is a field for '*开发者姓名' (Developer Name) with the value '王挺'. A note below says 'apikey (该apikey可以调用平台所有即用服务)'. The 'apikey' field contains the value '5341277306bb3825f19fa44f1318dade'.

2

A verification step for a phone number. It shows '*手机号码' (Phone Number) with the value '138****6645' and a button labeled '立即验证' (Verify Now).

3

Step 2: Visit the webpage of the API

eg: <http://apistore.baidu.com/apiworks/servicedetail/478.html>



全球天气
7天预报 · PM2.5 · 生活指数

即 V3

已调用9110万次 免费

| 访问 | 收藏 | 评论 |
|--------|-----|-----|
| 147316 | 564 | 227 |

中国和世界天气预报

服务商：和风全球天气 (企业)

所属分类：工具类、生活常用、服务类API、旅游、地图

更新时间：2016-05-09

apikey：获取apikey

服务简介：全球5万个城市（包含中国2567个市县地区）的7天天气预报、实况天气、每3小时天气、AQI空气质量、生活指数等数据，免费版有每天访问次数的限制，全能版可按需购买并增加了灾害预警等数据内容

<http://apistore.baidu.com/apiworks/servicedetail/880.html> 如果需要更多数据和更灵活的访问量请访问

<http://www.heweather.com>



Step 3: Parameter Adjusting

Your
API KEY

Parameters
Following
the URL,
which can
be input via
textbox on
the web
pages.

请求参数(header):

| 参数名 | 类型 | 必填 | 参数位置 | 描述 | 默认值 |
|--------|--------|----|--------|-------|------------|
| apikey | string | 是 | header | API密钥 | 您自己的apikey |

请求参数(urlParam):

| 参数名 | 类型 | 必填 | 参数位置 | 描述 | 默认值 |
|------|--------|----|----------|-----------------------|---------|
| city | string | 是 | urlParam | 城市名称, 国内城市支持中英文... 展开 | beijing |

请求示例:

```
curl示例  php示例  python示例  java示例  C#示例  ObjectC示例  Swift示例
1  # -*- coding: utf-8 -*-
2  import sys, urllib, urllib2, json
3
4  url = 'http://apis.baidu.com/heweather/weather/free?city=beijing'
5
6
7  req = urllib2.Request(url)
8
9  req.add_header("apikey", "您自己的apikey")
10
11  resp = urllib2.urlopen(req)
12  content = resp.read()
13  if (content):
14      print (content)
15
```

Tips: There would be some small modifications, if you use Python 3.

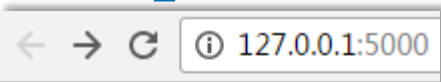


- *Step 4:*
Copy,
Paste,
and Revise
based on
“Login.py”

```
4 from flask import Flask
5 from flask import request
6 import urllib
7
8 app = Flask(__name__)
9
10 @app.route('/', methods=['GET', 'POST'])
11 def home():
12     return '''<h1>Home</h1>
13         <p><a href="/CityWeather">Weather Forecast</a></p>'''
14
15 @app.route('/CityWeather', methods=['GET'])
16 def login_form():
17     return '''<form action="/CityWeather" method="POST">
18         <p>City Name <input name="CityName"></p>
19         <p><button type="submit">Weather</button></p>
20         </form>'''
21
22 @app.route("/CityWeather", methods=['POST'])
23 def login():
24     # get data from Baidu:
25     url = 'http://apis.baidu.com/heweather/weather/free?city=' + request.form['CityName']
26     req = urllib.request.Request(url)
27     req.add_header("apikey", "5e7ef01f44164283876b0dd0cbd1461e") # Dont forget to change your spikey
28     resp = urllib.request.urlopen(req)
29     content = resp.read()
30     if (content):
31         return content
32     return '<h3>Bad City Name.</h3>'
33
34 if __name__ == '__main__':
35     app.run()
```

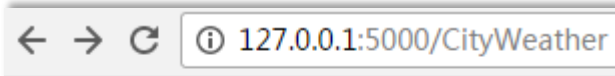


Step 5 : Testing



Home

[Weather Forecast](#)



City Name



```
127.0.0.1:5000/CityWeather
{"HeWeather data service 3.0":[{"aqi":{"city":"aqi":"47","co":"1","no2":"44","o3":"73","pm10":"34","pm25":"21","qlty":"优","so2":"15"},"basic":{"city":"上海","cnty":"中国","id":"CN101020100","lat":"31.213000","lon":"121.445000","update":{"loc":"2016-10-17 20:51","utc":"2016-10-17 12:51"},"daily_forecast":[{"astro":{"sr":"05:58","ss":"17:20"},"cond":{"code_d":"101","code_n":"104","txt_d":"多云","txt_n":"阴"},"date":"2016-10-17","hum":"81","pcpn":"0.0","pop":"0","pres":"1018","tmp":{"max":"25","min":"20"},"vis":"10","wind":{"deg":"34","dir":"东北风","sc":"微风","spd":"1"}},{"astro":{"sr":"05:59","ss":"17:19"},"cond":{"code_d":"305","code_n":"300","txt_d":"小雨","txt_n":"阵雨"},"date":"2016-10-18","hum":"84","pcpn":"0.2","pop":"89","pres":"1019","tmp":{"max":"24","min":"20"},"vis":"10","wind":{"deg":"68","dir":"东风","sc":"微风","spd":"0"}},{"astro":{"sr":"06:00","ss":"17:18"},"cond":{"code_d":"300","code_n":"104","txt_d":"阵雨","txt_n":"阴"},"date":"2016-10-19","hum":"86","pcpn":"0.1","pop":"9","pres":"1016","tmp":{"max":"24","min":"20"},"vis":"10","wind":{"deg":"78","dir":"东风","sc":"微风","spd":"8"}},{"astro":{"sr":"06:00","ss":"17:17"},"cond":{"code_d":"104","code_n":"104","txt_d":"阴","txt_n":"阴"},"date":"2016-10-20","hum":"84","pcpn":"0.1","pop":"40","pres":"1014","tmp":{"max":"26","min":"21"},"vis":"10","wind":{"deg":"82","dir":"东风","sc":"微风","spd":"5"}},{"astro":{"sr":"06:01","ss":"17:16"},"cond":{"code_d":"300","code_n":"300","txt_d":"阵雨","txt_n":"阵雨"},"date":"2016-10-21","hum":"90","pcpn":"6.7","pop":"61","pres":"1011","tmp":{"max":"24","min":"22"},"vis":"10","wind":{"deg":"94","dir":"东风","sc":"微风","spd":"9"}},{"astro":{"sr":"06:02","ss":"17:15"},"cond":{"code_d":"300","code_n":"307","txt_d":"阵雨","txt_n":"大雨"},"date":"2016-10-22","hum":"97","pcpn":"24.8","pop":"100","pres":"1006","tmp":{"max":"24","min":"20"},"vis":"9","wind":{"deg":"340","dir":"东北风","sc":"微风","spd":"7"}},{"astro":{"sr":"06:02","ss":"17:14"},"cond":{"code_d":"300","code_n":"104","txt_d":"阵雨","txt_n":"阴"},"date":"2016-10-23","hum":"90","pcpn":"37.8","pop":"100","pres":"1011","tmp":{"max":"23","min":"19"},"vis":"5","wind":{"deg":"44","dir":"东北风","sc":"微风","spd":"4"}],"hourly_forecast":[{"date":"2016-10-17 22:00","hum":"86","pop":"0","pres":"1019","tmp":"21","wind":{"deg":"51","dir":"东北风","sc":"微风","spd":"9"}],"now":{"cond":{"code":"101","txt":"多云"},"fl":"25","hum":"72","pcpn":"0","pres":"1019","tmp":"22","vis":"10","wind":{"deg":"60","dir":"东北风","sc":"3-4","spd":"16"},"status":"ok","suggestion":{"comf":{"brf":"舒适","txt":"白天不太热也不太冷,风力不大,相信您在这样的天气条件下,应会感到比较清爽和舒适。"},"cw":{"brf":"不宜","txt":"不宜洗车,未来24小时内有雨,如果在此期间洗车,雨水和路上的泥水可能会再次弄脏您的爱车。"},"drsg":{"brf":"舒适","txt":"建议着长袖T恤、衬衫加单裤等服装.年老体弱者宜着针织长袖衬衫、马甲和长裤。"},"flu":{"brf":"较易发","txt":"天气转凉,空气湿度较大,较易发生感冒,体质较弱的朋友请注意适当防护。"},"sport":{"brf":"较不宜","txt":"有降水,推荐您在室内进行健身休闲运动;若坚持户外运动,须注意携带雨具并注意避雨防滑。"},"trav":{"brf":"适宜","txt":"温度适宜,又有较弱降水和微风作伴,会给您的旅行带来意想不到的景象,适宜旅游,可不要错过机会呦!"},"uv":{"brf":"最弱","txt":"属弱紫外线辐射天气,无需特别防护.若长期在户外,建议涂擦SPF在8-12之间的防晒护肤品。"}]}]}}
```




Ask A Question



What are the advantages and disadvantages of using API?

•Advantages:

- much easier to get useful information*
- more efficient to develop website platforms*
- more helps from professionals*

•Disadvantages:

- maybe not free*
- maybe limitations*
- maybe some mistakes or exceptions*



human-readable text for data transmission

JSON

Analyzing JSON

1. `import json`
2. `json.loads`
3. `json.dumps`
4. get data following the tree
5. output results



JSON 结构视图

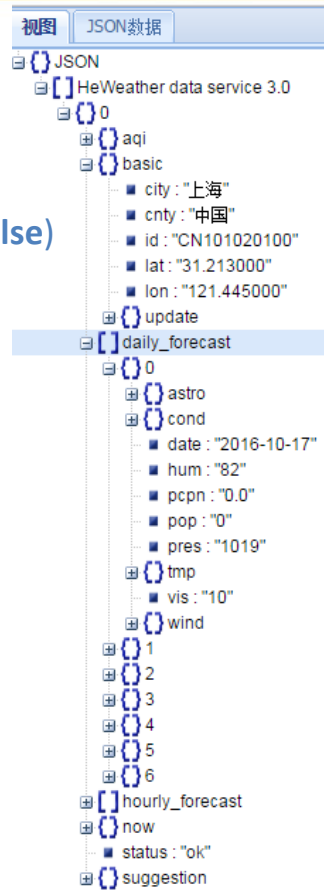
<http://www.bejson.com/jsonviewernew/>

```
city = json.dumps(response_json['HeWeather data service 3.0'][0]['basic'].get("city"),ensure_ascii=False)
```

有中文

```
date = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0].get("date"))
```

无中文



JSON

```
22 @app.route('/CityWeather', methods=['POST'])
23 def login():
24     # get data from Baidu:
25     url = 'http://apis.baidu.com/heweather/weather/free?city=' + request.form['CityName']
26     req = urllib.request.Request(url)
27     req.add_header("apikey", "5e7ef01f44164283876b0dd0cbd1461e") # Dont forget to change your apikey
28     resp = urllib.request.urlopen(req)
29     response = resp.read().decode('utf-8')
30     response_json = json.loads(response)
31     city = json.dumps(response_json['HeWeather data service 3.0'][0]['basic'].get("city"), ensure_ascii=False)
32     date = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0].get("date"))
33     max_tmp = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['tmp'].get("max"))
34     min_tmp = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['tmp'].get("min"))
35     txt_d = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['cond'].get("txt_d"),
36                       ensure_ascii=False)
37     txt_n = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['cond'].get("txt_n"),
38                       ensure_ascii=False)
39     wind_dir = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['wind'].get("dir"),
40                           ensure_ascii=False)
41     wind_sc = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['wind'].get("sc"),
42                           ensure_ascii=False)
43     wind_spd = json.dumps(response_json['HeWeather data service 3.0'][0]['daily_forecast'][0]['wind'].get("spd"))
44
45     content = city + "今日天气 (" + date + ") <br>" + "白天: " + txt_d + "<br>夜里: " + txt_n + \
46             "<br>最高气温: " + max_tmp + "<br>最低气温: " + min_tmp + "<br>风向: " + wind_dir + "<br>风速: " + wind_sc + "(" + wind_spd + "km/小时)"
47     if (content):
48         return content.replace("\n", "<br>")
49     return '<h3>Bad City Name. </h3>'
```





Example

← → ↻ ⓘ 127.0.0.1:5000/CityWeather

City Name

1



2

← → ↻ ⓘ 127.0.0.1:5000/CityWeather

北京今日天气 (2016-10-18)

白天: 晴

夜里: 霾

最高气温: 21

最低气温: 11

风向: 无持续风向

风速: 微风(4km/小时)



some examples on social media

Web API Applications

Wechat Selected

<http://apistore.baidu.com/apiworks/servicedetail/632.html>

API集市 > 生活常用 > 微信热门精选



The card features a speech bubble icon with three dots, representing the 'Wechat Selected' API. It includes a 'V4' badge, indicating the API version, and a '免费' (Free) label. The usage statistics show '已调用484万次' (Called 4.84 million times).

微信热门精选

服务商：天行数据 (个人)

所属分类：生活常用

更新时间：2016-06-28

apikey：获取apikey

服务简介：微信公众平台热门精选数据每小时自动更新一次，支持按时间、随机排序，支持指定来源和关键词检索。

访问

12943

收藏

234

评论

53



Web API Applications

Film Box Office

<https://www.juhe.cn/docs/api/id/44>



Juhe.cn
一家数据银行

首页

API

营销方案

聚合代码

新闻动态



电影票房

聚合数据

免费

支持HTTPS

电影票房, 票房榜

数据ID: 44

连接应用数: 5758

接口状态: 正常

接入服务商: 聚合数据

免费 / 无限次

申请数据



douban

https://developers.douban.com/wiki/?title=api_v2



豆瓣开发者服务

豆瓣开发者服务是基于豆瓣各产品、面向开发者的开放接口（API）服务。在这里，你可以接入豆瓣社区、读书、电影、音乐、同城等产品的优质内容，以及基于各种兴趣的用户关系。



在你的应用中使用豆瓣API

利用API快速构建Web、移动或桌面应用

用豆瓣帐号登录

快速积累注册用户，获取用户兴趣档案，让内容在豆瓣传播



上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

Homework

Homework

1. Each group should try your best to seek some APIs for data collection from the Internet. Approaches for using APIs can be found on each web site. Databases for storing those collected data should be designed before data collection according to different JSON structure. Programs showed in Lecture 4 and 5 can be integrated to solve the problem of both data collection from APIs and data storage.



Homework

2. We have learned how to get data from Wechat Selected, Film Box Office, and douban. Please select **ONE** of them and analyze the raw JSON data, then try to transfer some of them to a readable style. You may take **EXAMPLE 2: Weather Forecast 2** as an example. Students who submit this homework before the end of November will get at least 6 score (Max. 10) additionally. (optional)





上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

Reference

- XML
 - <http://www.w3school.com.cn/x.asp>
- JSON
 - <http://www.w3school.com.cn/json>



[美] Ryan Mitchell 著
陶俊杰 陈小莉 译

Reference

- Baidu (<http://apistore.baidu.com/>)
- Weibo (<http://open.weibo.com/wiki/微博API>)
- WeChat (<http://dev.wechat.com/>)
- QQ Tencent (<http://wiki.open.qq.com/wiki/API文档>)
- 360 (<http://open.360.cn/>)
- Taobao (http://open.taobao.com/docs/api_list.htm?spm=a219a.7386653.1.22.3XIaLs)
- JD (<http://jos.jd.com/>)
- Suning (<http://open.suning.com/>)
- Douban (<https://developers.douban.com/>)
- Juhe (<https://www.juhe.cn/>)
- Google (<https://developers.google.com/>)
- Facebook (<http://developers.facebook.com/>)
- Twitter (<https://dev.twitter.com/>)
- Linkedin (<https://developer.linkedin.com/>)





上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

The End of Lecture 5



Thank You



<http://www.wangting.ac.cn>