

Social Computing in Python for Rookies

Dr. Ting Wang's Online Courses

Dr. Ting Wang

SHANGHAI INTERNATIONAL STUDIES UNIVERSITY

Introduction to the Lecturer

Ting WANG (<http://www.wangting.ac.cn>)

- Associate Professor, Shanghai International Studies University
- Senior Engineer in Computer Science
- CIO, Shanghai nFitting Intelligent Science and Technology Co., Ltd.
- Academic Committee, Wuxi YOCSEF, CCF
- Postdoc, Tsinghua University, 2016
- PhD, University of Liverpool, 2013

Research Interests

- Artificial Intelligence
- Machine Learning
- Big Data
- Computer Vision
- Natural Language Processing

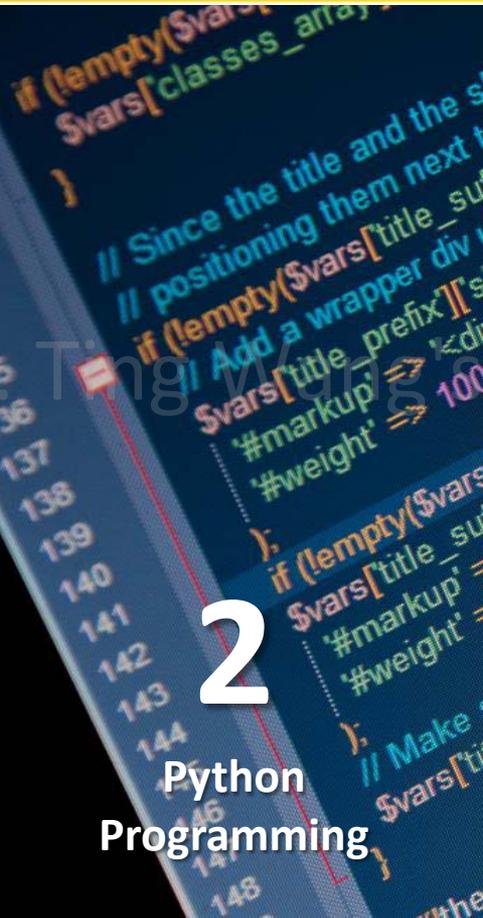


Outlines



1

Why Python?



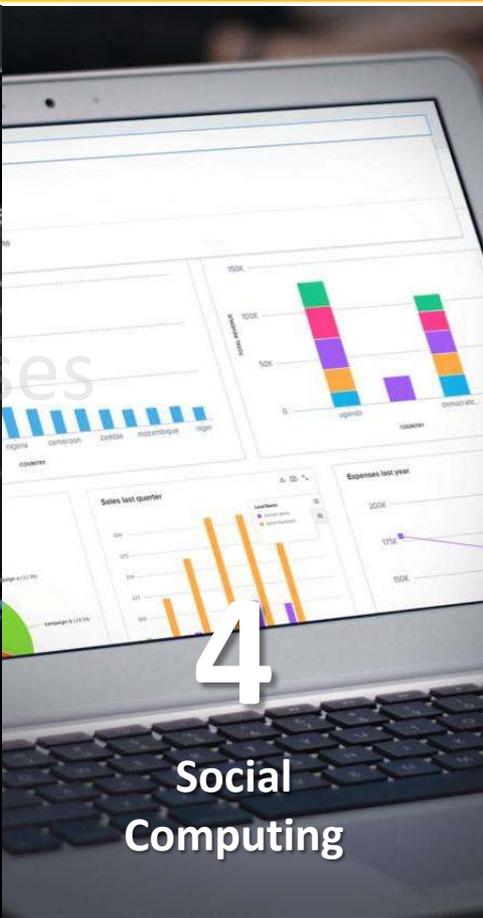
2

Python Programming



3

Files, DB and Web Developing



4

Social Computing

a widely used high-level programming language for general-purpose programming

Dr. Ting Wang's Online Courses

Why Python?

Why Choose Python?

The Language Python

Python is the king language
in the era of Big Data and AI.

It is:

- widely used
- easy to learn
- cross-platform
- powerful for engineering
- free to use
- well supported
- high paid

Language Rank	Types	Spectrum Ranking
1. C		100.0
2. Java		98.1
3. Python		98.0
4. C++		95.9
5. R		87.9
6. C#		86.7
7. PHP		82.8
8. JavaScript		82.2
9. Ruby		74.5
10. Go		71.9

<http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages>

What is Python?

The Birth of Python

Python is a widely used high-level, general-purpose, interpreted, dynamic programming language designed by Guido van Rossum in 1991.



Guido van Rossum

g Wang's Online Courses



Official Support

Official Website of Python

<https://www.python.org/>

The screenshot shows the Python.org homepage. At the top, there is a navigation bar with links for Python, PSF, Docs, PyPI, Jobs, and Community. Below this is the Python logo and a search bar with a 'GO' button. To the right of the search bar are links for 'Socialize' and 'Sign In'. A secondary navigation bar contains links for 'About', 'Downloads', 'Documentation', 'Community', 'Success Stories', 'News', and 'Events'. A prominent yellow button labeled '>_ Launch Interactive Shell' is centered below the navigation. The main content area features the text: 'Python is a programming language that lets you work quickly and integrate systems more effectively. >>> [Learn More](#)'. Below this, there are four columns of content: 'Get Started' (with a power icon), 'Download' (with a download icon), 'Docs' (with a book icon), and 'Jobs' (with a briefcase icon). Each column contains a brief description and a link to further resources.

Python

PSF

Docs

PyPI

Jobs

Community

python™

Search

GO

Socialize

Sign In

About

Downloads

Documentation

Community

Success Stories

News

Events

>_ Launch Interactive Shell

Python is a programming language that lets you work quickly and integrate systems more effectively. >>> [Learn More](#)

Get Started

Whether you're new to programming or an experienced developer, it's easy to learn and use Python.

Start with our [Beginner's Guide](#)

Download

Python source code and installers are available for download for all versions! Not sure which version to use? Check [here](#).

Latest: Python 3.6.1 - Python 2.7.13

Docs

Documentation for Python's standard library, along with tutorials and guides, are available online.

docs.python.org

Jobs

Looking for work or have a Python related position that you're trying to hire for? Our **relaunched community-run job board** is the place to go.

jobs.python.org

Code Support

GITHUB.COM

<https://github.com/>



Chinese Support

Liao Xue-Feng's Official Python Teaching Website

<http://www.liaoxuefeng.com/wiki/0014316089557264a6b348958f449949df42a6d3a2e542c000>

廖雪峰的官方网站

编程 读书 Java教程 **Python** JavaScript教程 Python教程 Git教程 问答 赞助

登录

目录

- Python教程
- Python简介
- 安装Python
 - Python解释器
- 第一个Python程序
 - 使用文本编辑器
 - Python代码运行助手
 - 输入和输出
- Python基础
 - 数据类型和变量
 - 字符串和编码
 - 使用list和tuple
 - 条件判断
 - 循环
 - 使用dict和set

Python教程 2.7旧版教程

阅读: 5090797

这是小白的Python新手教程，具有如下特点：

中文，免费，零起点，完整示例，基于最新的Python 3版本。

Python是一种计算机程序设计语言。你可能已经听说过很多种流行的编程语言，比如非常难学的C语言，非常流行的Java语言，适合初学者的Basic语言，适合网页编程的JavaScript语言等等。

那Python是一种什么语言？

首先，我们普及一下编程语言的基础知识。用任何编程语言来开发程序，都是为了让计算机干活，比如下载一个MP3，编写一个文档等等，而计算机干活的CPU只认识机器指令，所以，尽管不同的编程语言差异极大，最后都得“翻译”成CPU可以执行的机器指令。而不同的编程语言，干同一个活，编写的代码量，差距也很大。

比如，完成同一个任务，C语言要写1000行代码，Java只需要写100行，而Python可能只要20行。

所以Python是一种相当高级的语言。

你也许会问，代码少还不好？代码少的代价是运行速度慢，C程序运行1秒钟，Java程序可能需要2秒，而Python程序可能就需要10秒。

Installation

1. Python Installation Package

Python 3.X

<https://www.python.org/downloads/>

2. Integrated Development Environment

Pycharm

<http://www.jetbrains.com/pycharm/>

3. Python Package Management

easy_install

https://pypi.python.org/pypi/ez_setup

pip

<https://pypi.python.org/pypi/pip>

4. Third Party Packages

Unofficial Packages

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

Install
Python

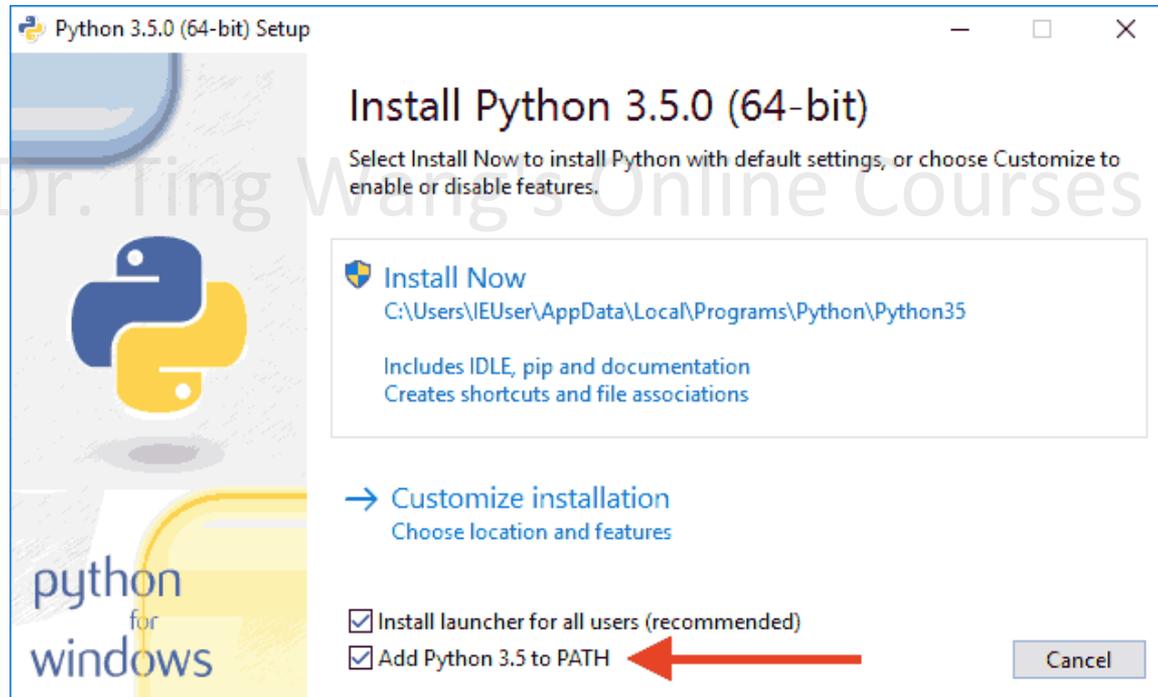
Dr. Ting Wang's Online Courses

Installation

1. Python Installation Package

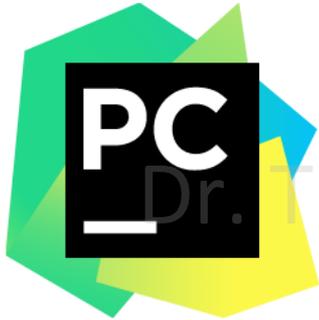
<https://www.python.org/downloads/windows/>

Download Windows x86-64 executable installer



2. Integrated Development Environment

<http://www.jetbrains.com/pycharm/download/#section=windows>



Version: 2017.1.2

Build: 171.4249.47

Released: April 28, 2017

[System requirements](#)

[Installation Instructions](#)

[Previous versions](#) [↗](#)

Download PyCharm

Windows

macOS

Linux

Professional

Full-featured IDE
for Python & Web
development

DOWNLOAD

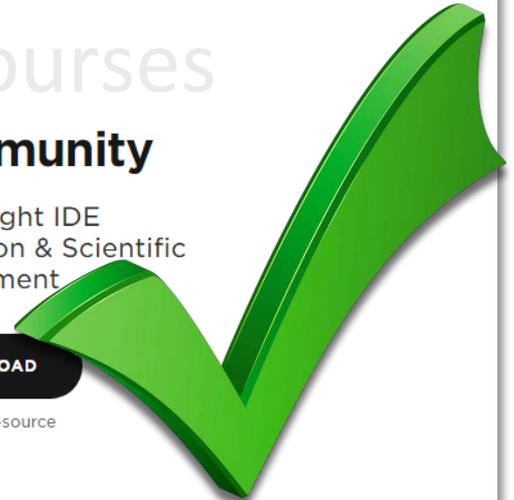
Free trial

Community

Lightweight IDE
for Python & Scientific
development

DOWNLOAD

Free, open-source



3. Python Package Management

“easy_install” and “pip” are used for installing packages from PyPI, a public Python resource.

easy_install

Download : https://pypi.python.org/pypi/ez_setup

Installation : `python ez_setup.py`

To install a package:

```
easy_install <package_name>
```

```
easy_install "<package_name>==<version>"
```

To upgrade a package:

```
easy_install -U "<package_name>>=<version>"
```

pip

Download : <https://pypi.python.org/pypi/pip>

Installation : `python setup.py install`

To install a package:

```
pip install <package_name>
```

```
pip install <package_name>==<version>
```

To upgrade a package:

```
pip install --upgrade <package_name>>=<version>
```

To uninstall a package:

```
pip uninstall <package_name>
```

4. Third Party Packages

Another Choice: Unofficial Packages

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

Unofficial Windows Binaries for Python Extension Packages

by **Christoph Gohlke**, Laboratory for Fluorescence Dynamics, University of California, Irvine.

This page provides 32- and 64-bit Windows binaries of many scientific open-source extension packages for the official [CPython distribution](#) of the [Python](#) programming language.

The files are unofficial (meaning: informal, unrecognized, personal, unsupported, no warranty, no liability, provided "as is") and made available for testing and evaluation purposes.

If downloads fail reload this page, enable JavaScript, disable download managers, disable proxies, clear cache, and use Firefox. Please only download files manually as needed.

Most binaries are built from source code found on [PyPI](#) or in the projects public revision control systems. Source code changes, if any, have been submitted to the project maintainers or are included in the packages.

Refer to the documentation of the individual packages for license restrictions and dependencies.

Use [pip](#) version 8 or newer to [install the downloaded .whl files](#). This page is not a pip package index.

Many binaries depend on [numpy-1.11+mkl](#) and the Microsoft Visual C++ 2008 ([x64](#), [x86](#), and [SP1](#) for CPython 2.7), Visual C++ 2010 ([x64](#), [x86](#), for CPython 3.4), or the Visual C++ 2015 ([x64 and x86](#) for CPython 3.5 and 3.6) redistributable packages.

Install [numpy+mkl](#) before other packages that depend on it.

The binaries are compatible with the most recent official CPython distributions on Windows >=6.0. Chances are they do not work with custom Python distributions included with Blender, Maya, ArcGIS, OSGeo4W, ABAQUS, Cygwin, Pythonxy, Canopy, EPD, Anaconda, WinPython etc. Many binaries are not compatible with Windows XP or Wine.

The packages are ZIP or 7z files, which allows for manual or scripted installation or repackaging of the content.

The files are provided "as is" without warranty or support of any kind. The entire risk as to the quality and performance is with you.

The opinions or statements expressed on this page should not be taken as a position or endorsement of the Laboratory for Fluorescence Dynamics or the University of California.

how to use Python and write code in Python

Dr. Ting Wang's Online Courses

Python Programming

Python Programming

**EXAMPLE 1:
My First Program**



My First Program

STEP 1

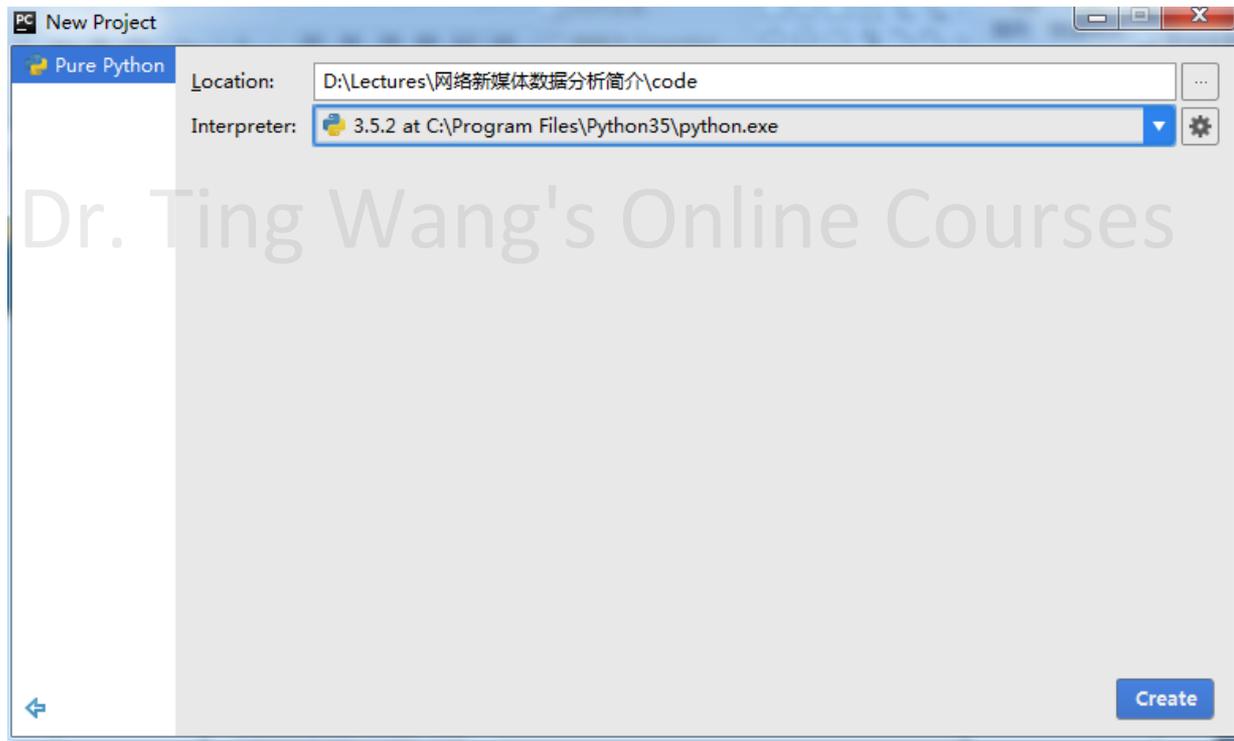
Create New Project



My First Program

STEP 2

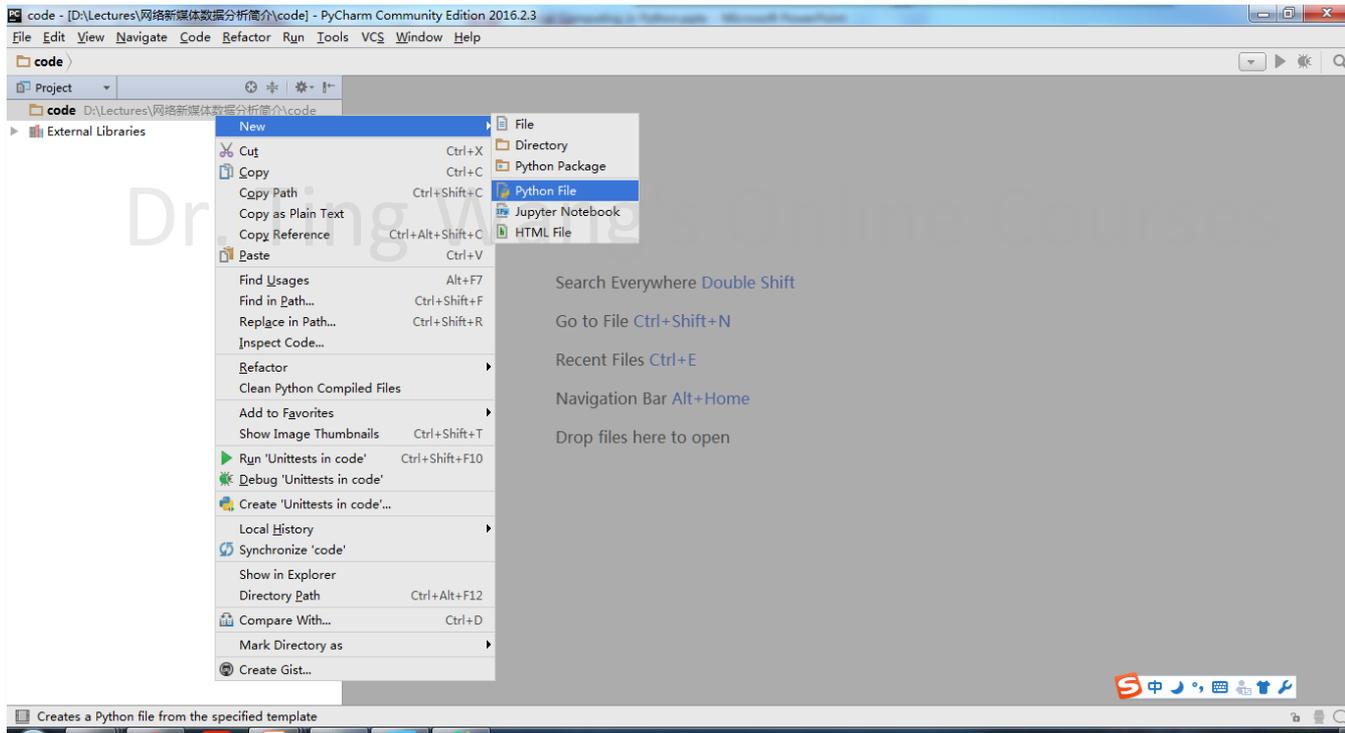
Select the location to save the code



My First Program

STEP 3

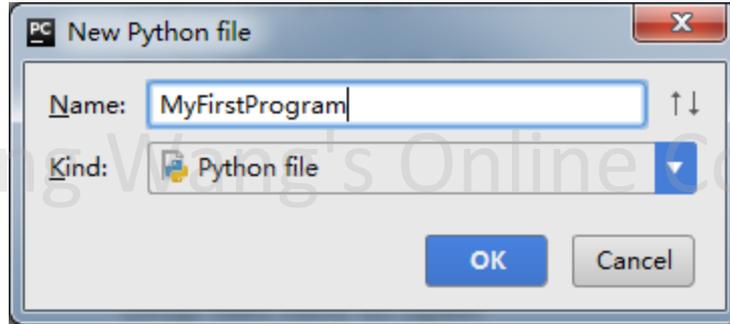
Right click on the folder to set up a new Python file



My First Program

STEP 4

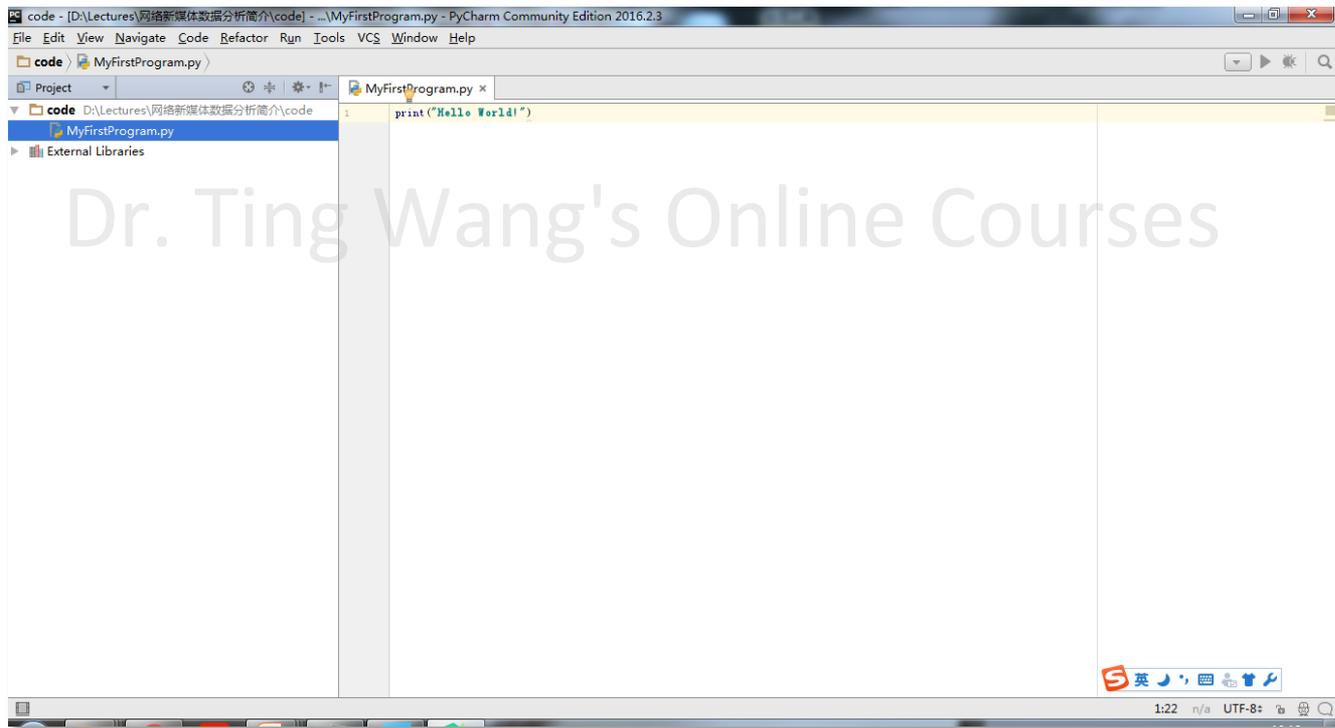
Give a name to your first Python file



My First Program

STEP 5

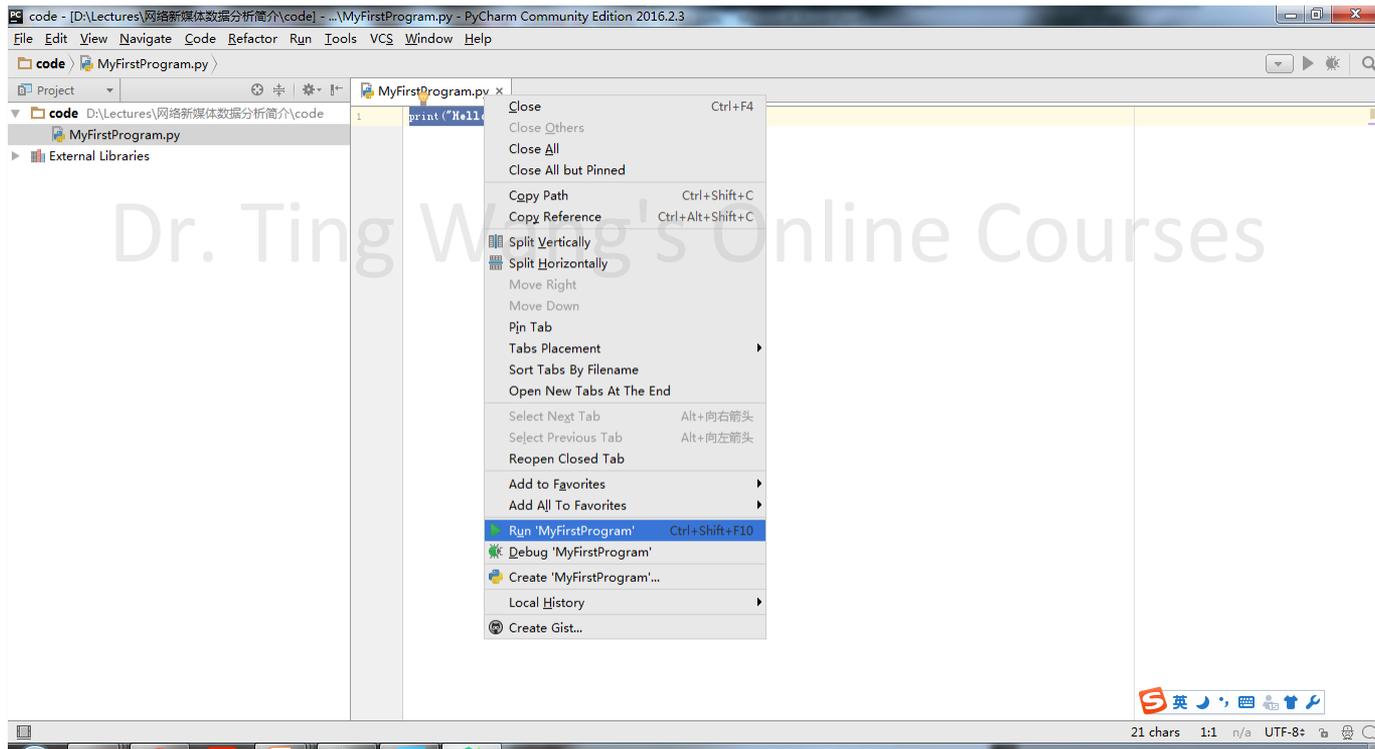
Put the code “`print("Hello World!")`” into the right large box.



My First Program

STEP 6

Right click on the tab and run it!



My First Program

STEP 6

The result



The screenshot shows a 'Run' window titled 'MyFirstProgram'. The command executed is `"C:\Program Files\Python35\python.exe" D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py`. The output is `Hello World!` followed by `Process finished with exit code 0`. A large watermark 'Dr. Ting Wang's Online Courses' is overlaid on the image.

```
Run MyFirstProgram
"C:\Program Files\Python35\python.exe" D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py
Hello World!
Process finished with exit code 0
```

My First Program

**Practice
Makes
Perfect**

1. Put the code

```
print(200+300)
```

into the Pycharm, run it, and check the result.

2. Compare the code and results with the following code

```
print("200+300")
```

and

```
print(Hello World!)
```

Dr. Ting Wang's Online Courses

Variable and its Value

a **storage location** paired with an associated symbolic **name** (an identifier), which contains some known or unknown quantity of information referred to as a **value**.

```
x=2  
name="Thomas"
```

Python is case sensitive.

They are different variables.

```
x=2  
X=3
```

Data Structure of Variables

#	Type	#	Type
1	Number	5	List
2	Bool	6	Tuple
3	None	7	Dict
4	String	8	Datetime

Reference: <http://www.cnblogs.com/linjiqin/p/3608541.html>

Grammar

Number

Integer

```
x=2
```

Float

```
y=2.0
```

Bool

True, False, and, or, not

```
x=2  
y=2.0  
print(x>y)  
print(x==y)  
print(not(x>y))
```

None

```
x=None  
print(x)
```

String

```
name="Thomas"
```

EXAMPLE 2:
input and output

OUTPUT

INPUT

Input and Output

Put the following code into the Pycharm, run it, and check the result.

```
name=input("What is your name?\n")
age=input("How old are you?\n")
print("Hello %, you are %d years old" %(name, int(age)))
```

Special Characters in Python:

Wildcard Character

%d
%f
%s

Convert

int()
str()

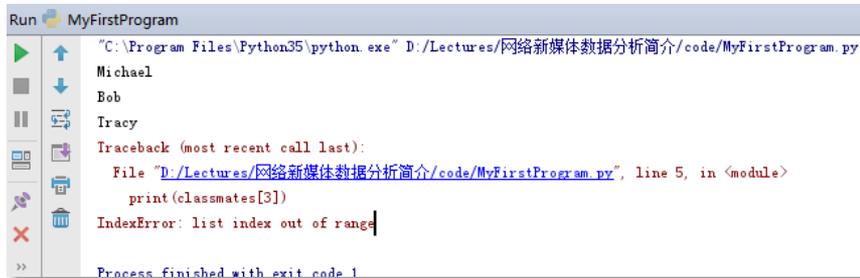
ESC, EEscape Character

For this	Use this	Setting x to:	Printing x will yield:
'	\'	'Don\'t do that'	Don't do that
"	\"	"She said \"hi\""	She said "hi"
\	\\	"Backslash: \\"	Backslash: \
[newline]	\n	"1\n2"	1 2
[carriage return]	\r	"1\r2"	2 overwrites the 1
[horizontal tab]	\t	"1\t2"	1 2
[backspace]	\b	"12\b3"	13
[16 bit unicode]	\uxxxx	"上海: \u4e0a\u6d77"	上海: 上海

List

- use “[]”
- start from 0

```
classmates = ['Michael', 'Bob', 'Tracy']  
print(classmates[0])  
print(classmates[1])  
print(classmates[2])  
print(classmates[3])
```



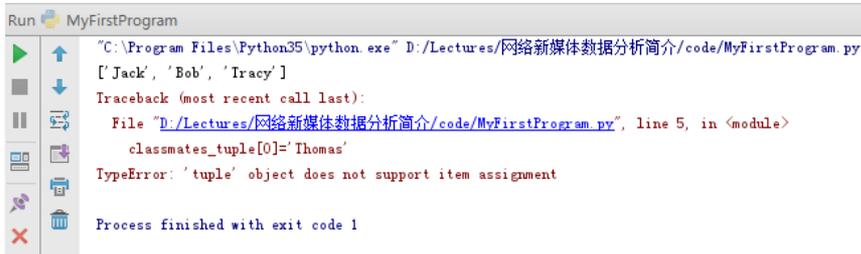
Run MyFirstProgram

```
"C:\Program Files\Python35\python.exe" D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py  
Michael  
Bob  
Tracy  
Traceback (most recent call last):  
  File "D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py", line 5, in <module>  
    print(classmates[3])  
IndexError: list index out of range  
Process finished with exit code 1
```

Tuple

- use “()”
- A list where values CANNOT be changed.

```
classmates_list = ['Michael', 'Bob', 'Tracy']  
classmates_tuple = ('Michael', 'Bob', 'Tracy')  
classmates_list[0]='Jack'  
print(classmates_list)  
classmates_tuple[0]='Thomas'  
print(classmates_tuple)
```



Run MyFirstProgram

```
"C:\Program Files\Python35\python.exe" D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py  
['Jack', 'Bob', 'Tracy']  
Traceback (most recent call last):  
  File "D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py", line 5, in <module>  
    classmates_tuple[0]='Thomas'  
TypeError: 'tuple' object does not support item assignment  
Process finished with exit code 1
```

Dict

- use {}
- `d = {'key1':value1, 'key2':value2}`
- Repetitions will be discarded

```
d = {'Michael': 95, 'Bob': 75, 'Tracy': 85, 'Tracy': 65}
print(d['Michael'])
print(d)
```

Set

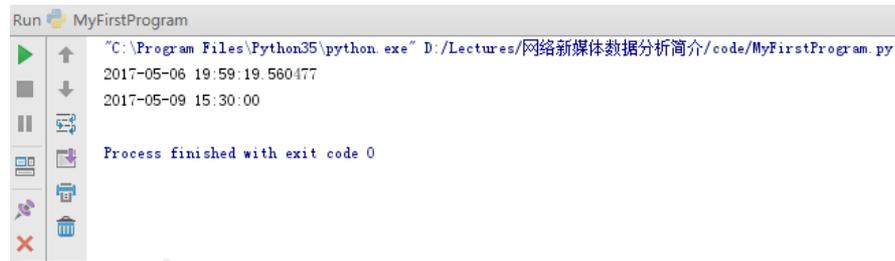
- use “set(())”
- `s = set([key1, key2, key3])`
- Repetitions will be discarded
- No value, only key
- Actually, set is a function

```
s = set([1,1,2,2,3,3])
print(s)
```

Datetime

```
import datetime
print(datetime.datetime.now())

# 用指定日期时间创建datetime
dt = datetime.datetime (2017, 5, 9, 15, 30)
print(dt)
```



```
Run MyFirstProgram
"C:\Program Files\Python35\python.exe" D:/Lectures/网络新媒体数据分析简介/code/MyFirstProgram.py
2017-05-06 19:59:19.560477
2017-05-09 15:30:00
Process finished with exit code 0
```

Note:

1. “datetime” is a module. It should be imported before it is employed.
2. Python has many modules for different usages. Moreover, there are also a great number of third-party modules, which can be installed by Python command “pip”.

Conditional Statement

```
if <condition 1>:  
    <statement1>  
elif < condition 2>:  
    <statement2>  
elif < condition 3>:  
    <statement3>  
else:  
    <statement4>
```

Iteration

```
for <counter> in <range>:  
    <statement>
```

```
while <condition>:  
    <statement>
```

break: stop the whole iteration

continue: stop this round, but continue to start the next round of this iteration

Question: What is the result of the following code?

```
for i in range(100):  
    if i%2==0:  
        print(i)
```

Function

Define Functions

```
def FunctionName(parameter1, parameter2,...)
    <statement>
    [return value]
```

← Optional

Call Functions

```
FunctionName(para1, para2,...)
```

```
def EvenNum(num):
    for i in range(num):
        if i%2==0:
            print(i)

number=input("Please input the range:\n")
EvenNum(int(number))
```

try...except...finally...

If we are not sure whether there are some errors in our code, we can use this statement.

Step 1: “try”

Step 2: Errors occur, stop “try”;

Step 3: go to “except”, and finish this part

Step 4: if there is a “finally” part then execute it;

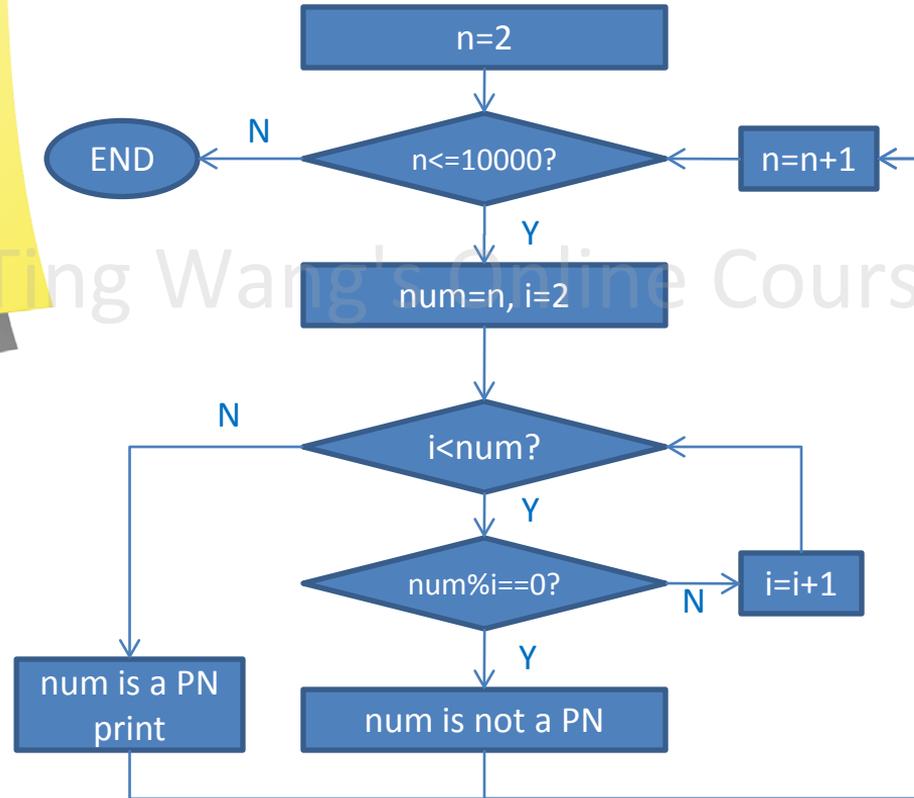
Step 5: finish

```
try:
    print('try...')
    r = 10 / 0
    print('result:', r)
except ZeroDivisionError as e:
    print('except:', e)
finally:
    print('finally...')
print('END')
```

Grammar

Practice
Makes
Perfect

1. Print all the Prime Numbers (PN) smaller than 10,000.



```
def PrimeNumber(num):
```

```
    i = 2
```

```
    while i < num:
```

```
        if num % i == 0:
```

```
            break;
```

```
            i = i + 1
```

```
        else:
```

```
            print(num)
```

```
number = 100
```

```
n = 2
```

```
while n <= number:
```

```
    PrimeNumber(n)
```

```
    n = n + 1
```

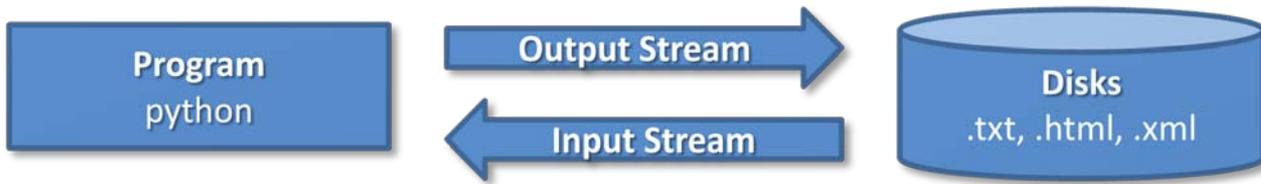
advanced python programming

Files, DB and Web Developing

Dr. Ting Wang's Online Courses

Files and Stream I/O

Stream Input / Output



Read Files (Input)

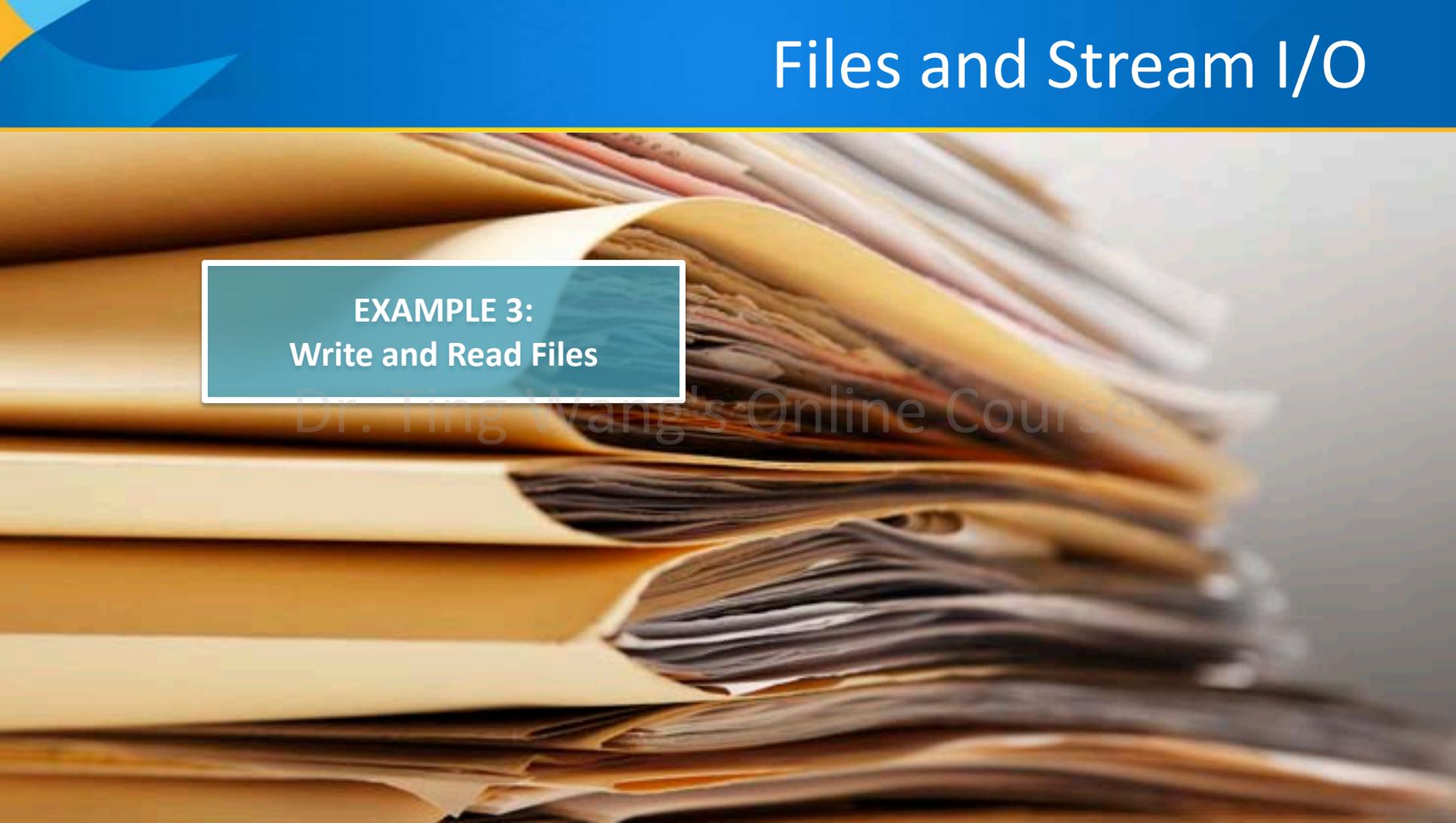
```
with open('FILE_PATH', 'r/rb') as VARIABLE_NAME  
VARIABLE_NAME.read()
```

Write Files (Output)

```
with open('FILE_PATH', 'w/wb') as VARIABLE_NAME  
VARIABLE_NAME.write('TEXT')
```

[Notes] Absolute and Relative Paths

Files and Stream I/O



**EXAMPLE 3:
Write and Read Files**

Dr. Jing Wang's Online Courses

Files and Stream I/O

Document Write-Save-Read

```
import datetime
with open('test.txt', 'w') as f:
    f.write('Today is')
    f.write(datetime.datetime.now().strftime('%Y-%m-%d'))

with open('test.txt', 'r') as f:
    s = f.read()
    print('open for read...')
    print(s)
```

[Notes]

If there is no such a file,
create it!

otherwise,
rewrite it!

Dr. Ting Wang's Online Courses

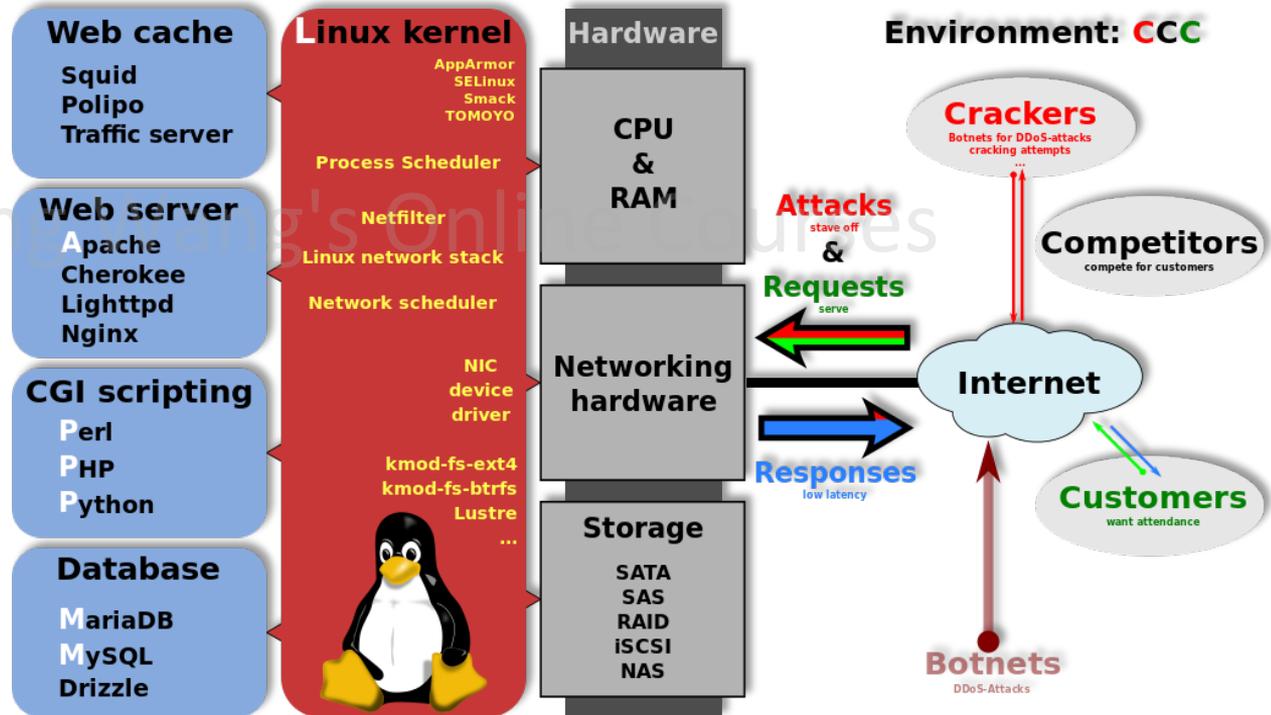
Data Base

MySQL

an open-source relational database management system (RDBMS)

LAMP

- Linux
- Apache
- MySQL
- Perl+PHP+Python



Installation

Official Website: <https://www.mysql.com/>

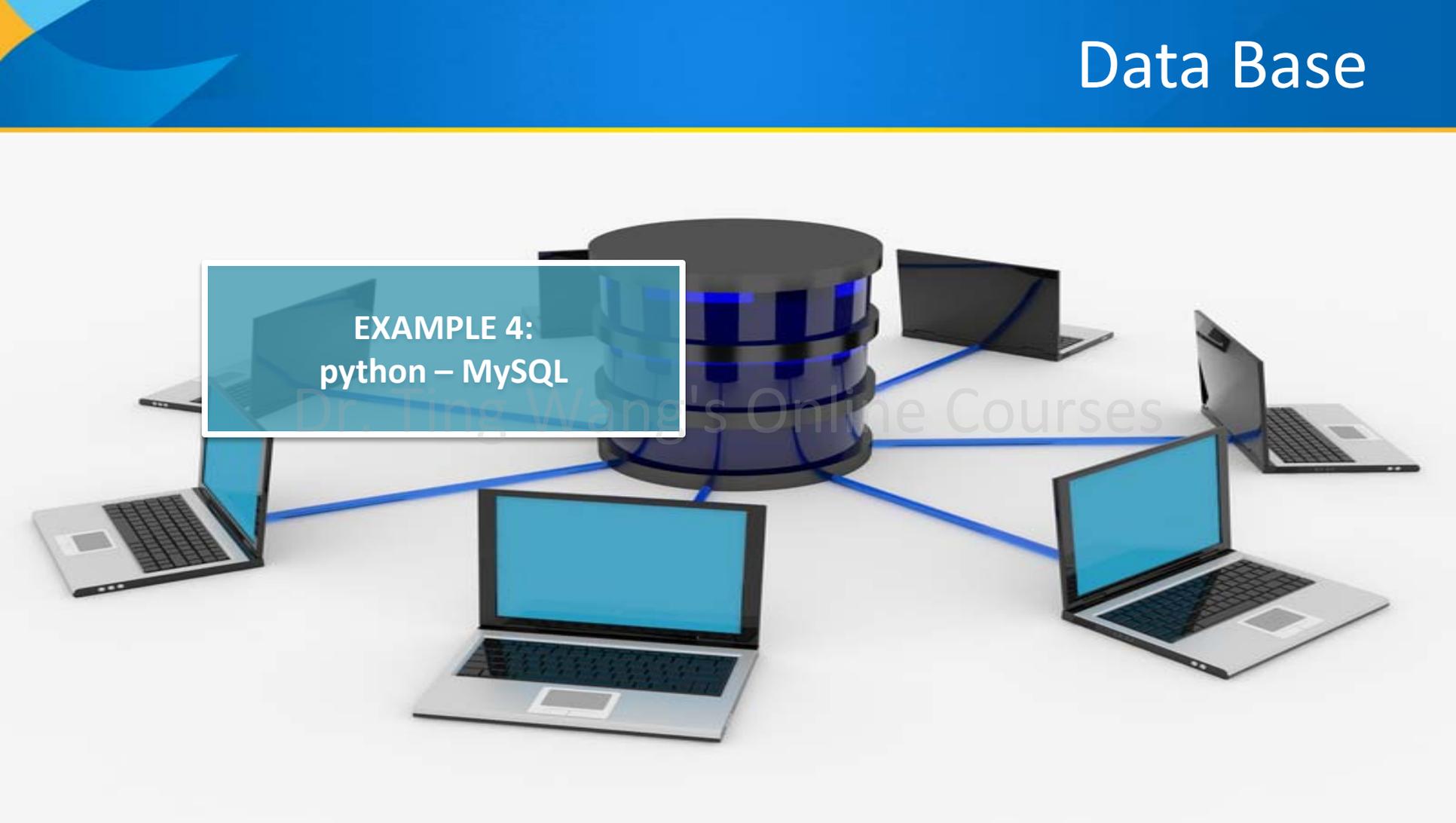
Download: MySQL Community Edition



pymysql

a third-party module for the connection between Python and MySQL

```
pip install pymysql
```

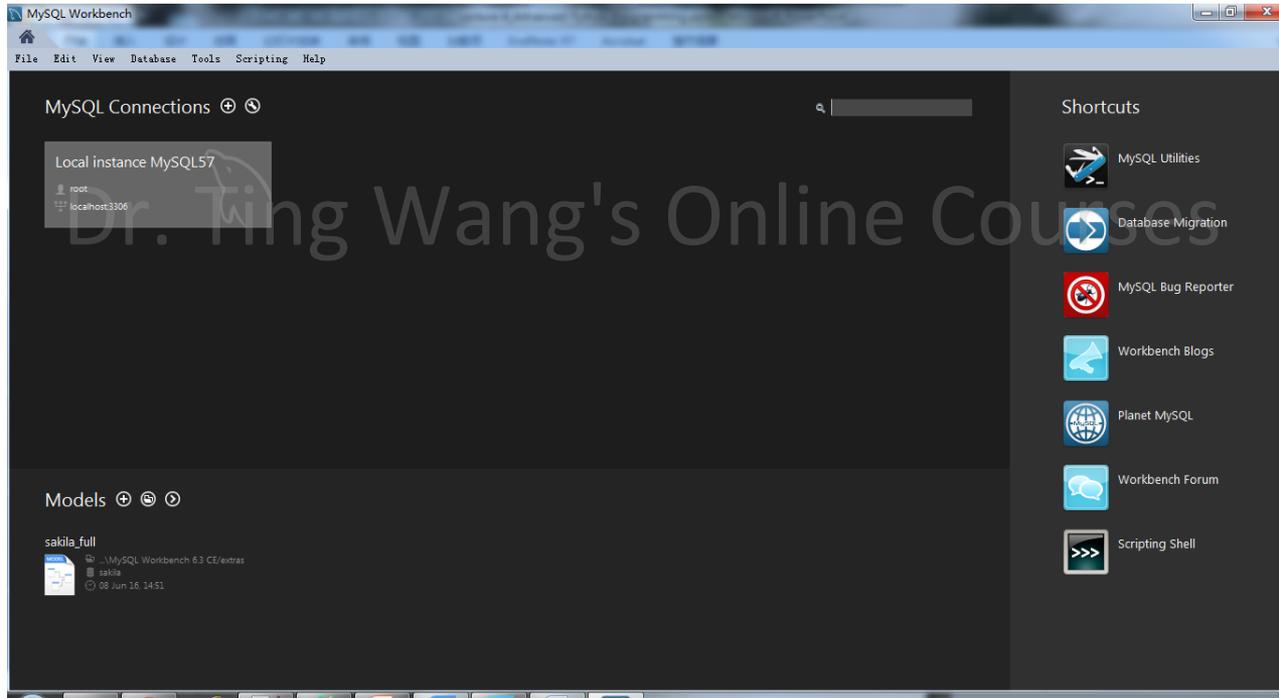


EXAMPLE 4:
python – MySQL

Dr. Ting Wang's Online Courses

Connect to Database

Start MySQL Workbench



Connect to Database

Create Tables

1. Create a new schema 

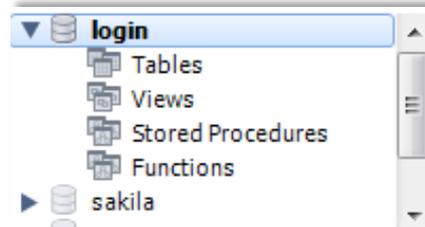
2. Set as a default schema (right click)

3. Run CreateMySQL.sql  

```
CREATE TABLE CUSTOMER (  
  CUSTOMER_ID      INT(10) PRIMARY KEY AUTO_INCREMENT,  
  CUSTOMER_NAME    VARCHAR(100) NOT NULL,  
  CUSTOMER_PWD     VARCHAR(100) NOT NULL);
```

4. Run InsertMySQL.sql

```
INSERT INTO CUSTOMER(CUSTOMER_NAME,CUSTOMER_PWD) VALUES('Thomas Edison','sisu');  
INSERT INTO CUSTOMER(CUSTOMER_NAME,CUSTOMER_PWD) VALUES('Albert Einstein','sisu');  
INSERT INTO CUSTOMER(CUSTOMER_NAME,CUSTOMER_PWD) VALUES('Marie Curie','shisu');
```



5. Result

```
select * from customer
```

CUSTOMER_ID	CUSTOMER_NAME	CUSTOMER_PWD
1	Thomas Edison	sisu
2	Albert Einstein	sisu
3	Marie Curie	shisu
NULL	NULL	NULL

Code

```
import pymysql
```

```
#1.Connection Open
```

```
conn = pymysql.connect(user='root',  
password='123456', database='login')
```

```
#2.Cursor Creating:
```

```
cursor = conn.cursor()
```

```
#3.SQL Execution
```

```
sqlstr = 'select * from customer where  
customer_name="Thomas Edison" and  
customer_pwd="shisu"'
```

```
#4.Cursor Moving
```

```
cursor.execute(sqlstr)
```

```
numrows = len(cursor.fetchall())
```

```
conn.commit()
```

```
#5.Connection Close
```

```
cursor.close()
```

```
conn.close()
```

```
# get data from request:
```

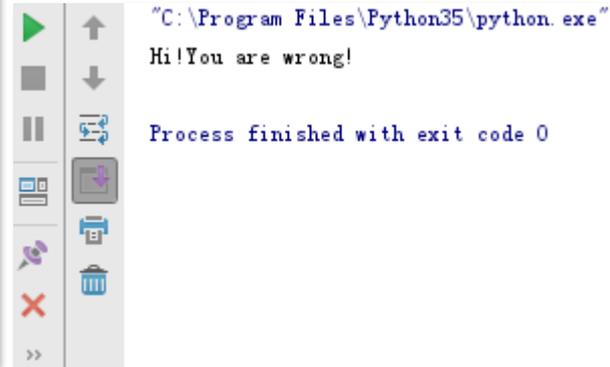
```
if numrows>0:
```

```
    print('Hello, Thomas Edison! ')
```

```
else:
```

```
    print('Hi!You are wrong! ')
```

Result:



```
"C:\Program Files\Python35\python.exe"  
Hi!You are wrong!  
  
Process finished with exit code 0
```

Flask (web framework)

a web framework for the Python programming language

Installation

```
pip install flask
```

Other Python

Web Frameworks

- Django
- Web.py
- Tornado
- Bottle

```
D:\SISU\新媒体数据分析与应用\Lecture 4\Flask>pip install flask
Collecting flask
  Using cached Flask-0.11.1-py2.py3-none-any.whl
Collecting Jinja2>=2.4 (from flask)
  Using cached Jinja2-2.8-py2.py3-none-any.whl
Collecting click>=2.0 (from flask)
  Using cached click-6.6.tar.gz
Collecting Werkzeug>=0.7 (from flask)
  Using cached Werkzeug-0.11.11-py2.py3-none-any.whl
Collecting itsdangerous>=0.21 (from flask)
  Using cached itsdangerous-0.24.tar.gz
Collecting MarkupSafe (from Jinja2)=2.4->flask)
  Using cached MarkupSafe-0.23.tar.gz
Installing collected packages: MarkupSafe, Jinja2, click, Werkzeug, itsdangerous, flask
  Running setup.py install for MarkupSafe ... done
  Running setup.py install for click ... done
  Running setup.py install for itsdangerous ... done
Successfully installed Jinja2-2.8 MarkupSafe-0.23 Werkzeug-0.11.11 click-6.6 flask-0.11.1 itsdangerous-0.24
```

EXAMPLE 5: Flask Login



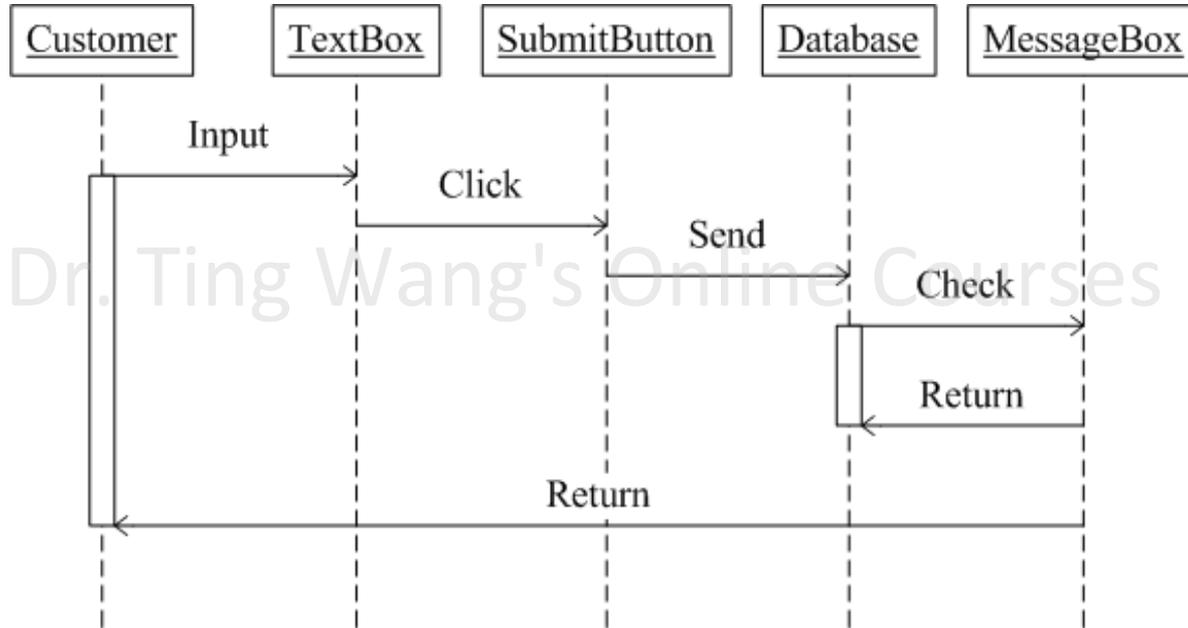
Adam .

FIN

Sign-in options

Web Programming

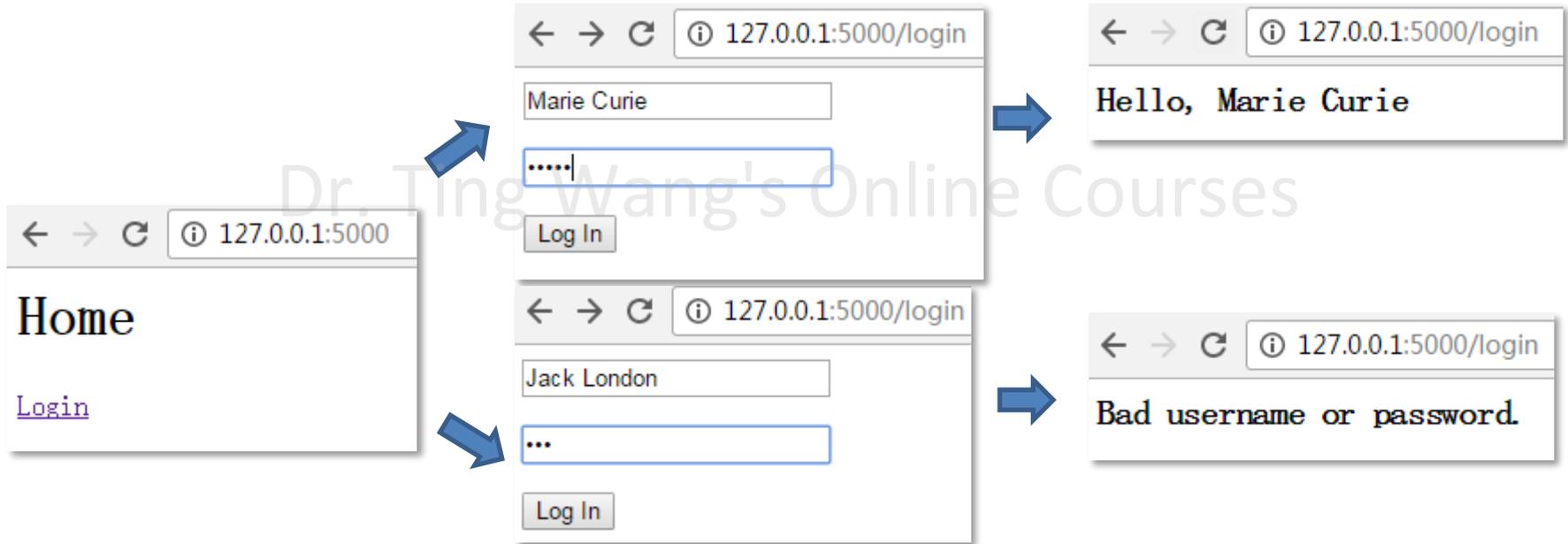
Login



Dr. Ting Wang's Online Courses

Web Programming

Testing

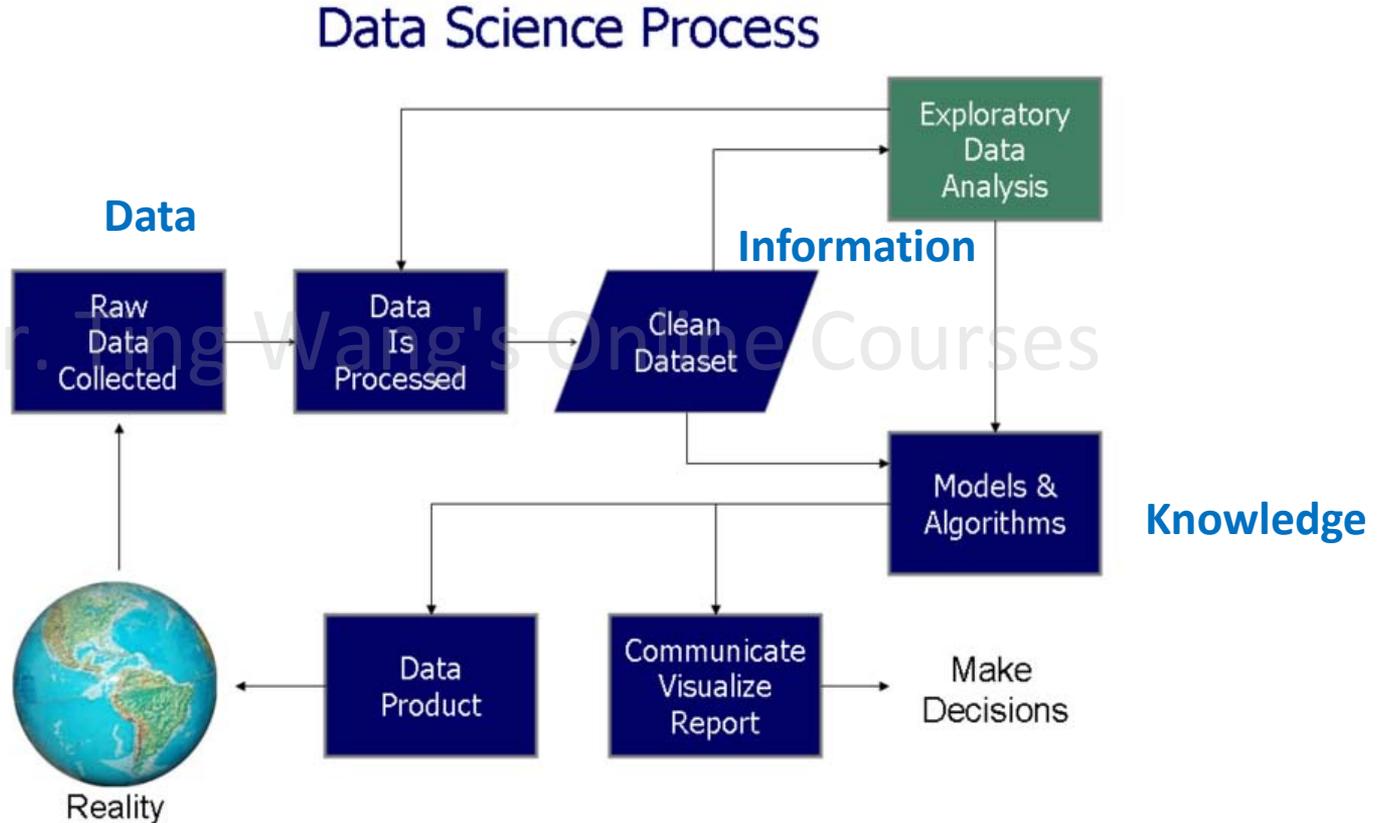


process of sampling signals for real world measurement with digital numeric values

Dr. Ting Wang's Online Courses

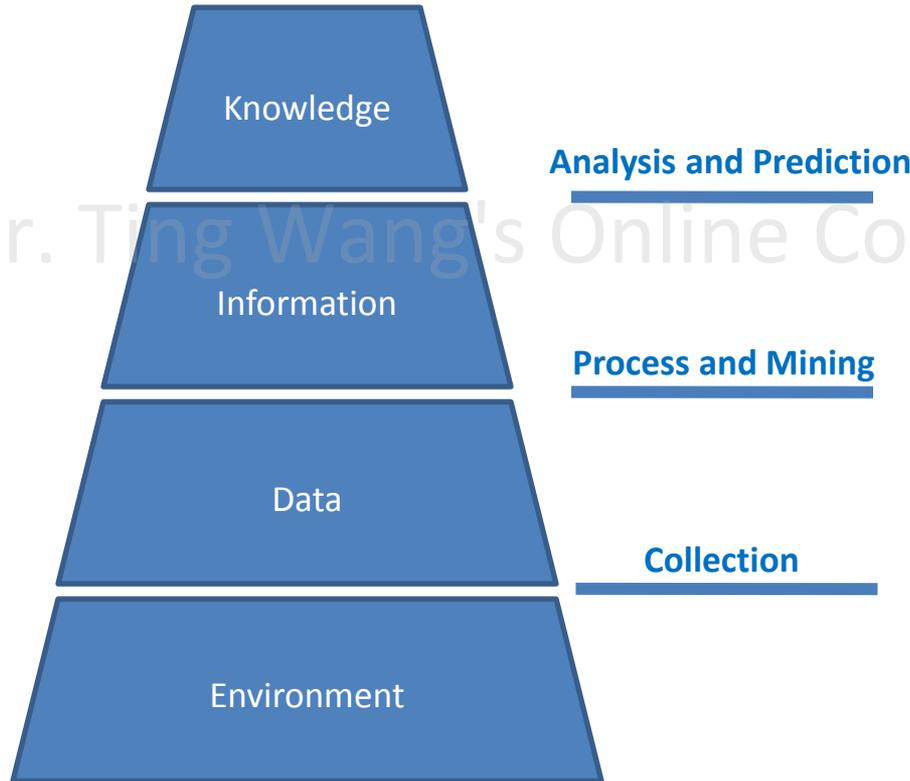
Social Computing

Data Science Process



Data Science Process

Relationship between data, information and knowledge



Dr. Ting Wang's Online Courses

Data Science Process

Methodology

1

Information
Acquisition

Web Data



2

Data Cleaning and
Information Retrieval

*Structured
Web
Information*



3

Knowledge Fusion and
Information Updating

*Feature
Analysis
and
Updating*



4

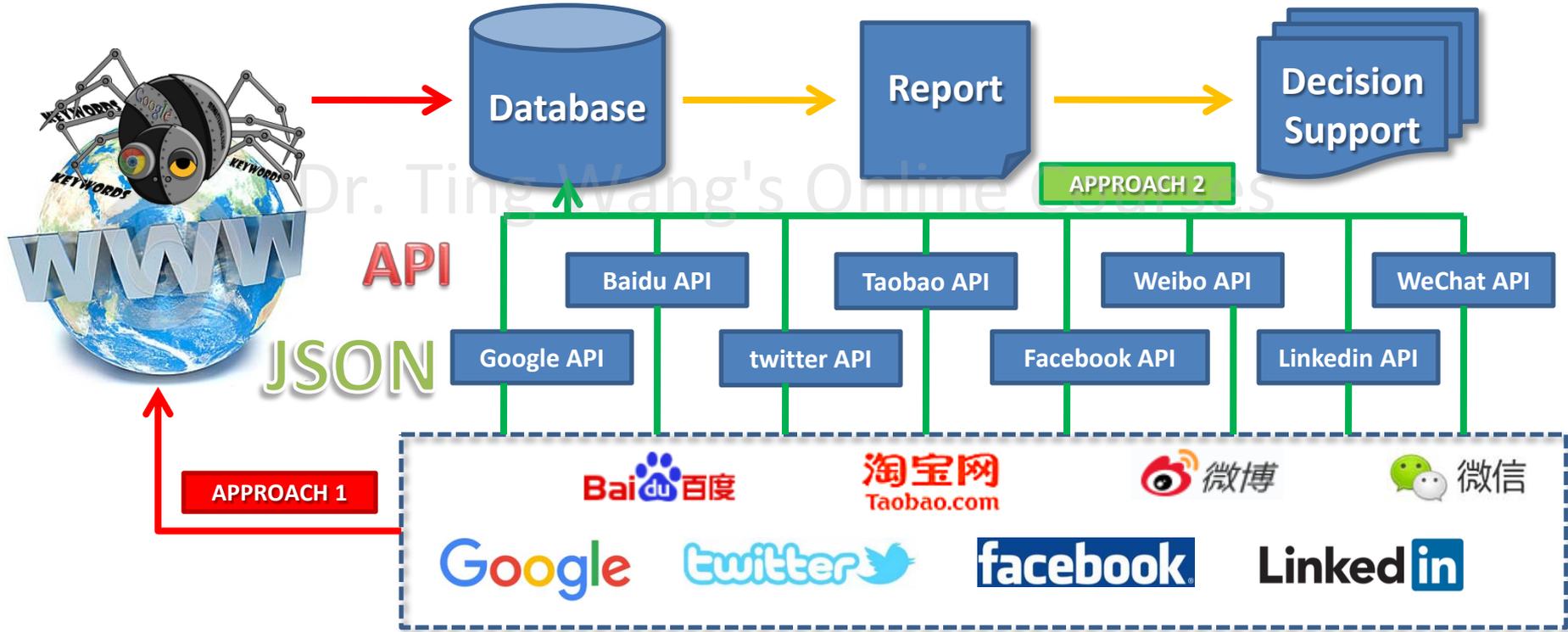
Prediction or Classification

*Personalized
Products or
Services*



Information Acquisition and Data

How to get data from social media?



Data Collection with Web Crawler

Web Crawler

A very simple example:

```
import urllib.request
response = urllib.request.urlopen('http://www.entgroup.cn/news/Markets/0941733.shtml')
HTMLText = response.read()

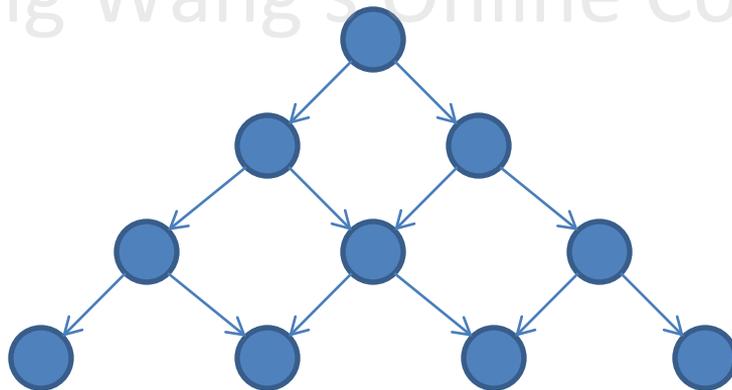
with open('webcrawler.html', 'wb') as f:
    f.write(HTMLText)
```

Data Collection with Web Crawler

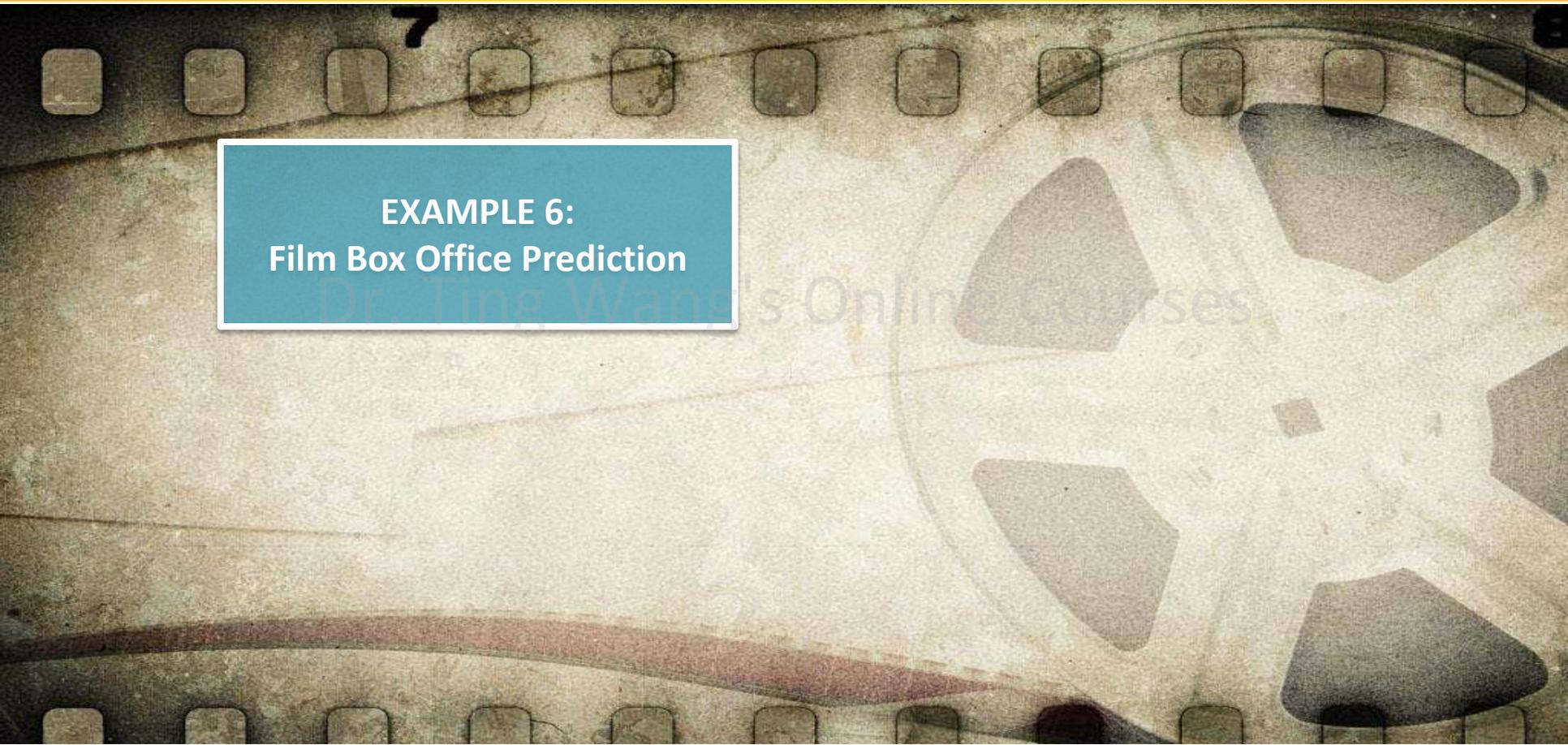
Deciding What to Search

1. URL list for the websites you want to search
2. Do nothing but search web pages via hyperlinks one by one
3. Depth-First-Search (DFS) and Breadth-First-Search (BFS)

Dr. Ting Wang's Online Courses



Data Collection with Web Crawler



**EXAMPLE 6:
Film Box Office Prediction**

Dr. Ting Wang's Online Courses

Film Box Office Prediction

Problem Description

Film Box Office Prediction

- is crucial to film investment
- is significant to the market with out Completion Bond
- can be done by a number of approaches

In this case, film box office prediction will be computed based on the information collected by online film news reports.

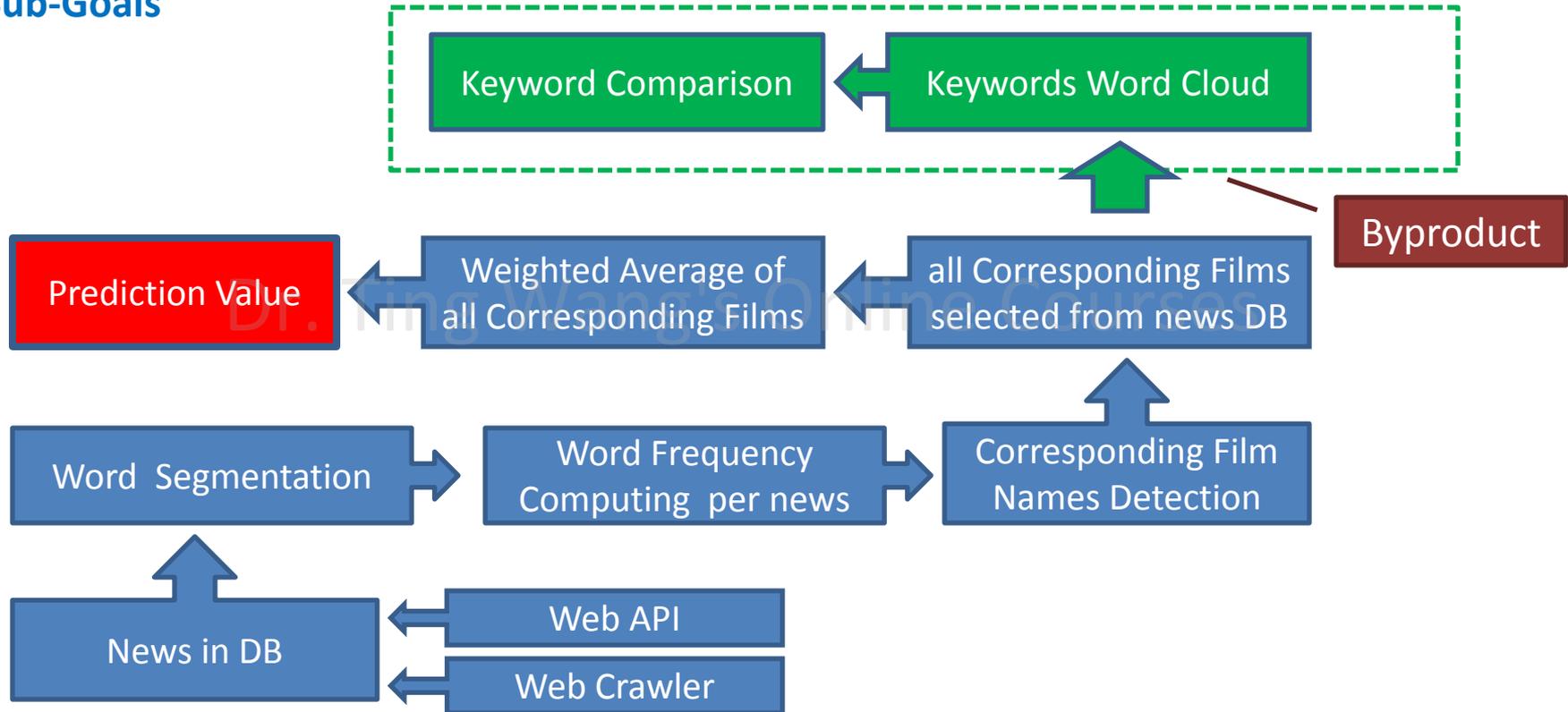
Terminal Goal

To make a decision: whether a film is worth of being invested or not.

This depends on the prediction value of the box office of the potential film project.

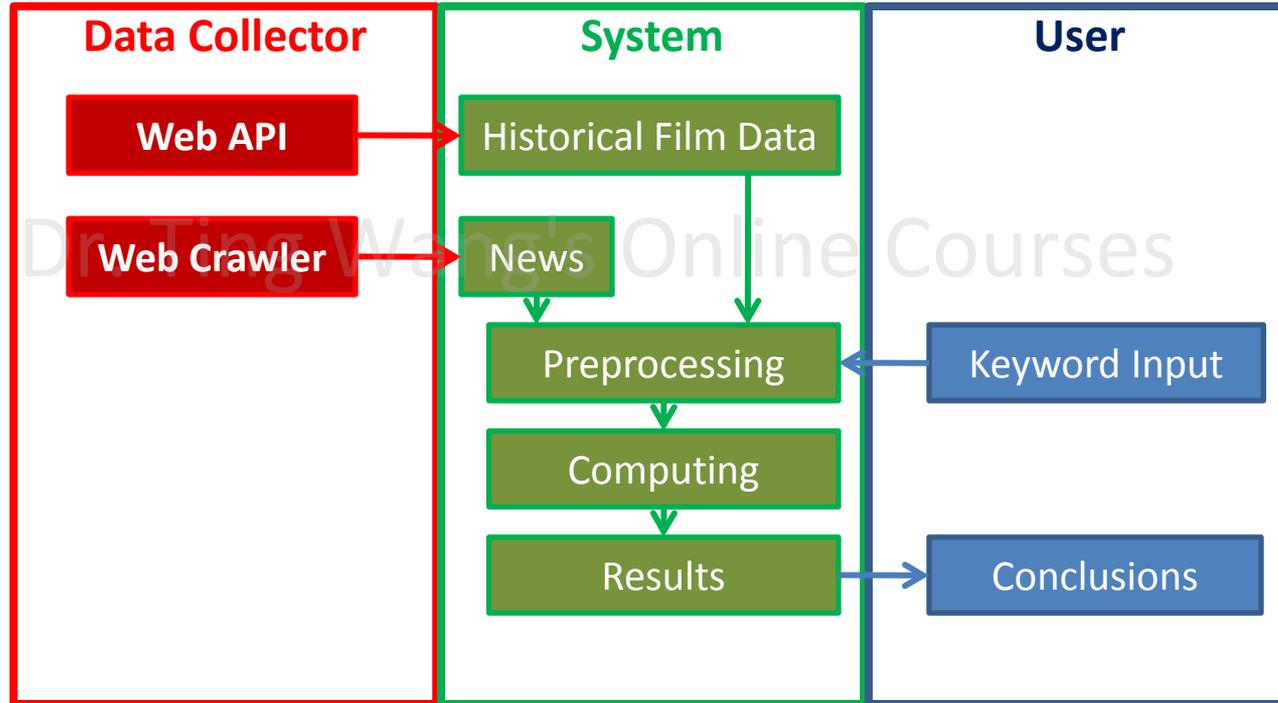
Film Box Office Prediction

Sub-Goals



Film Box Office Prediction

Activity Diagram



Film Box Office Prediction

Functions

1. Film Box Office Prediction
2. Byproduct: Keyword Comparison
 - Word Cloud
 - Media Attention
 - Feature Comparisons

Input and Output

Input: Keywords of film name

- Byproduct: Keywords
- Other conditions: Word Frequency, Periods,...

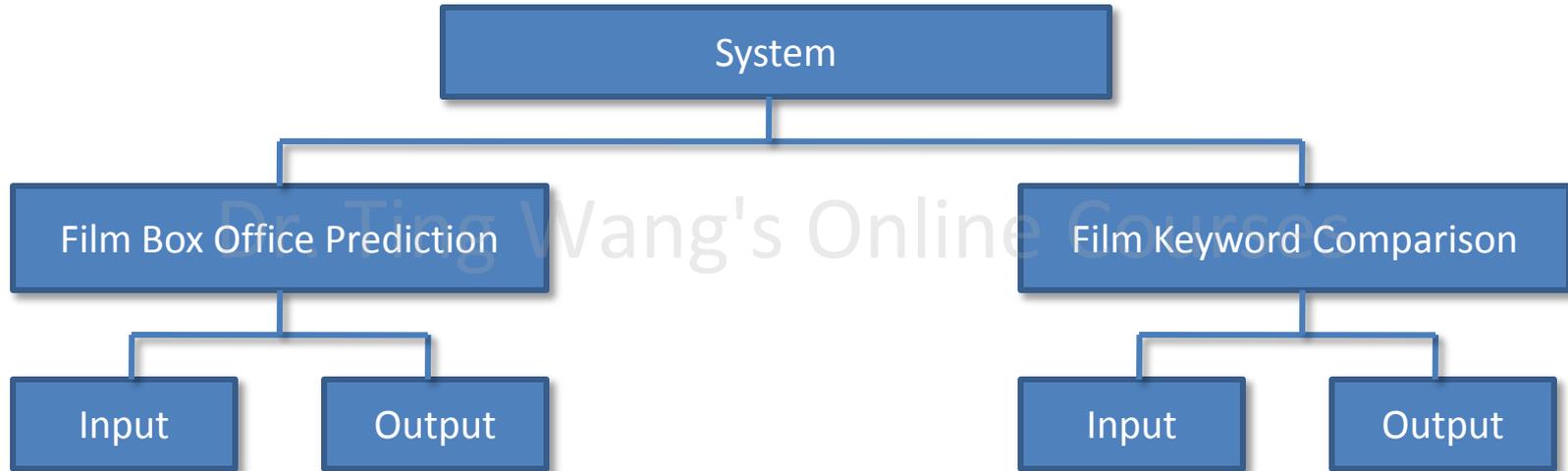
Output: Prediction value of film box office

- Word Cloud,
- Media Attention,
- Word Frequency Comparison

Dr. Ting Wang's Online Courses

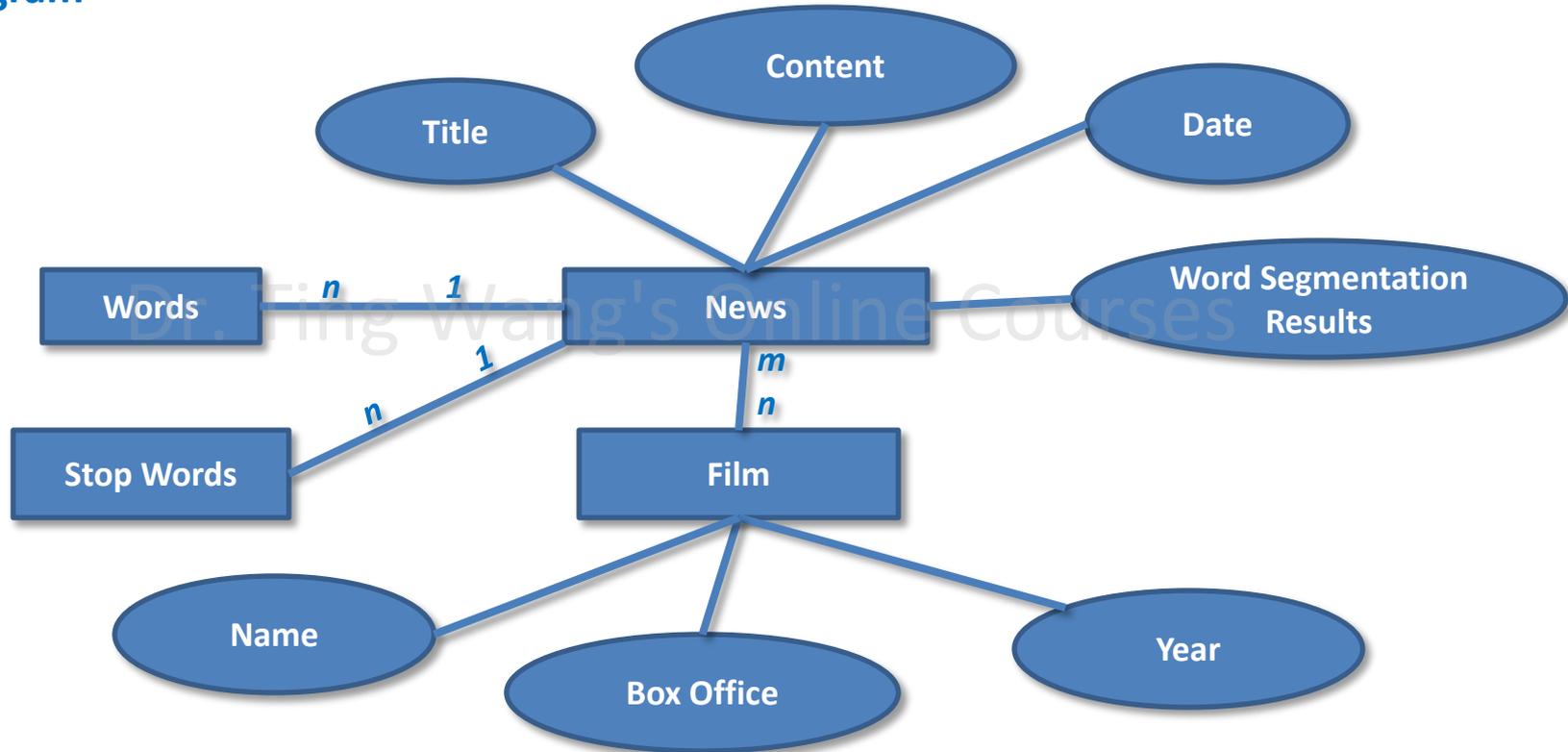
Film Box Office Prediction

System Design



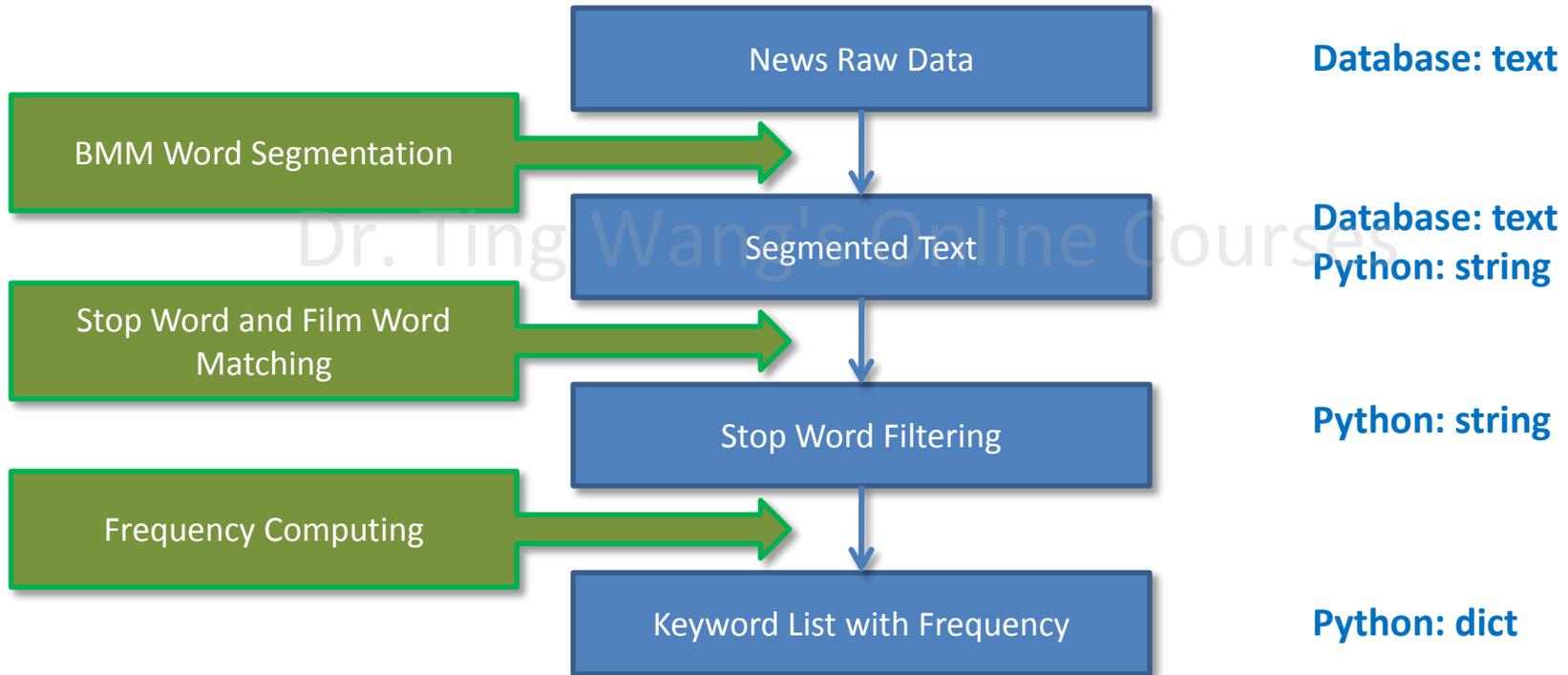
Film Box Office Prediction

ER Diagram



Film Box Office Prediction

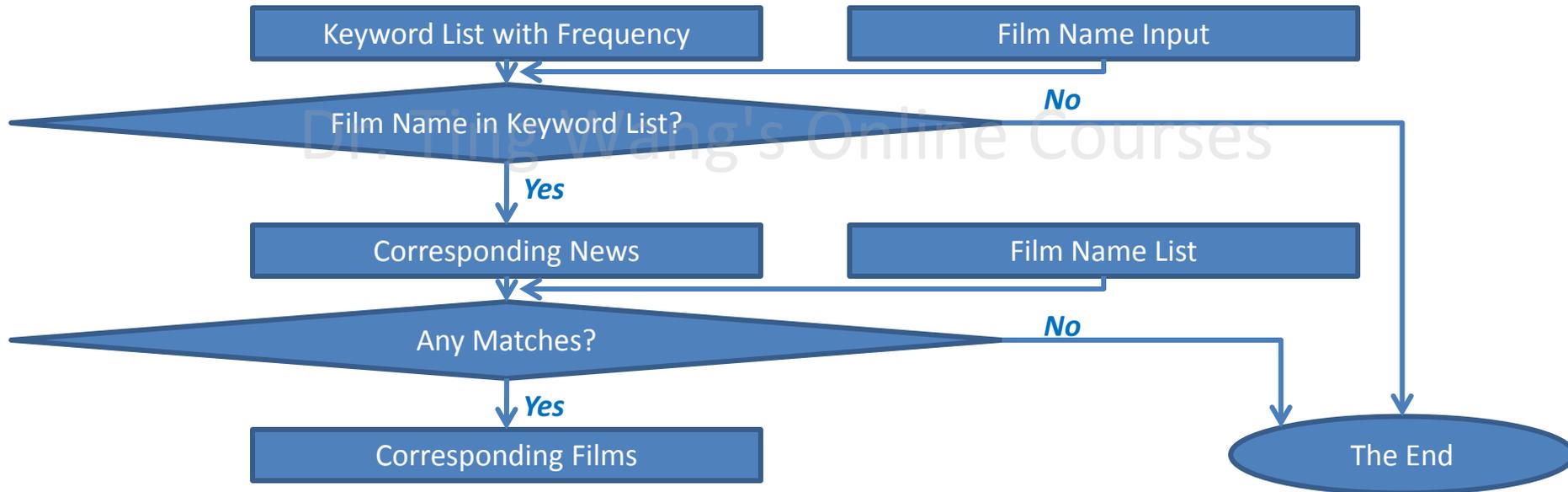
Data Transformation



Film Box Office Prediction

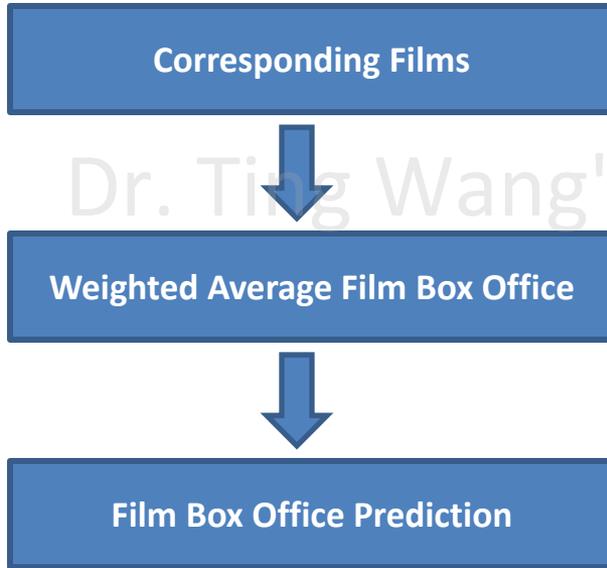
Information Acquisition (From Data to Info.)

For Film Box Office Prediction



Film Box Office Prediction

Prediction and Data Visualization



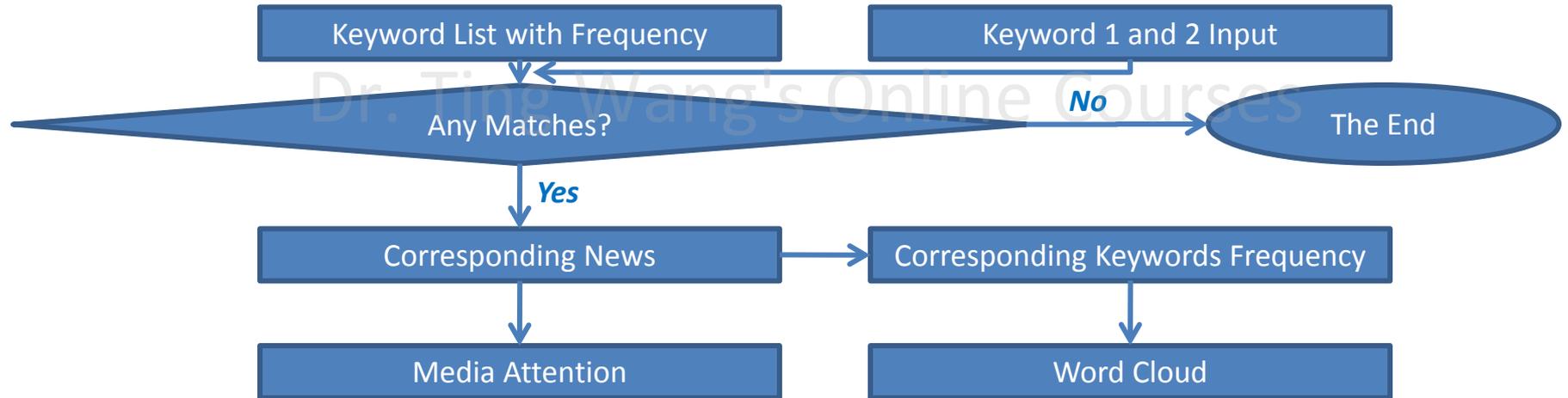
Dr. Ting Wang's Online Courses

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{n}$$

Film Box Office Prediction

Text Mining

For Byproduct, Keyword Comparison



Film Box Office Prediction

Software Development

- Python
- PyCharm
- Flask
- MySql

Data Source



News, 24100 articles

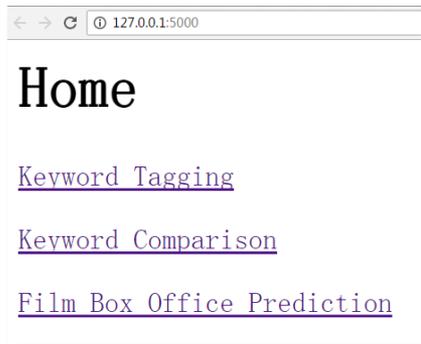
<http://www.entgroup.cn/>
All the news from 2007.11-2016.11

Films, 1893 movies

<http://58921.com/>
All the data of film box office
from 2008.1-2016.11

Dictionary, 149921 words and 508 stop words

Testing



Film Box Office Prediction

Film Box Office Prediction

The Great Wall (2016)

[Home](#)

Film Box Office of 长城 : 116401.04395604396(x10,000) RMB

Real Box Office:

1.173 Billion RMB

Predictive Box Office:

1.164 Billion RMB

Parameter Settings:

Please input the Film Name:

长城

Please input the Frequency of Keyword:

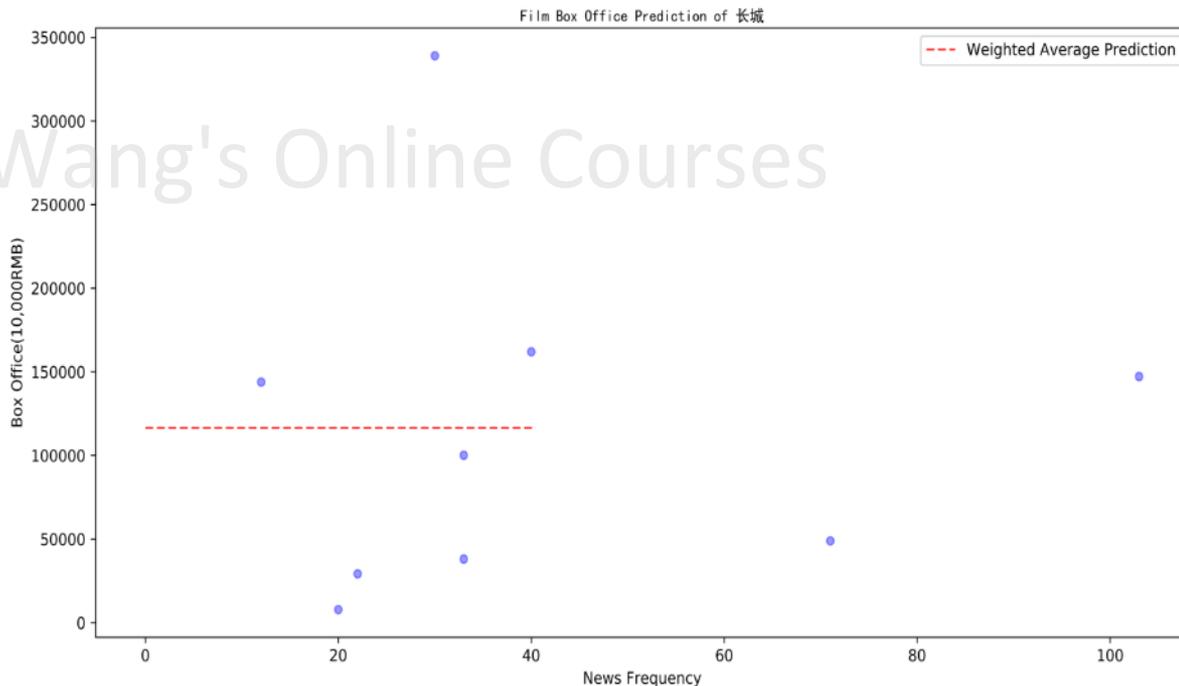
10

Start Date - End Date of News

2015-1-1

2016-12-1

Predict

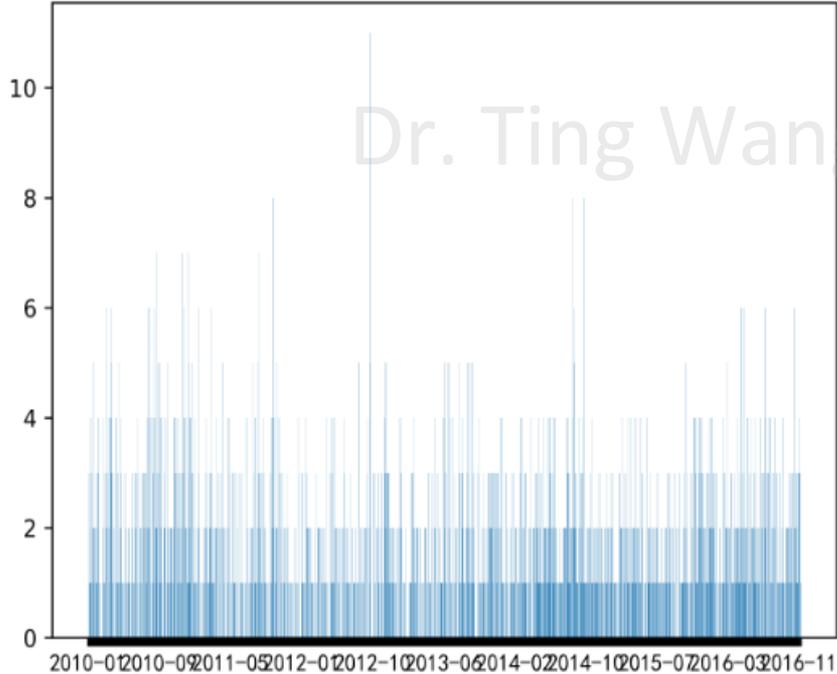


Film Box Office Prediction

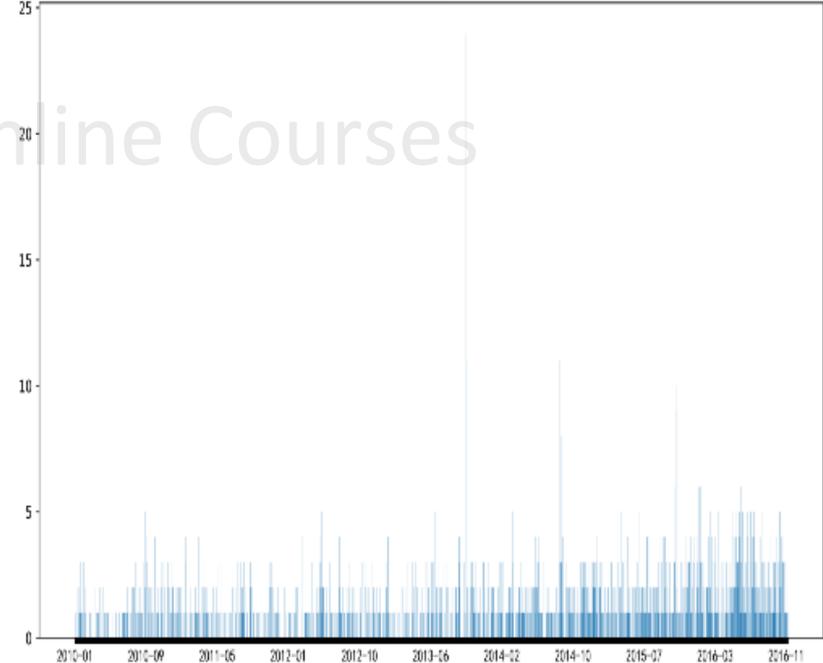
Media Attention



媒体热度

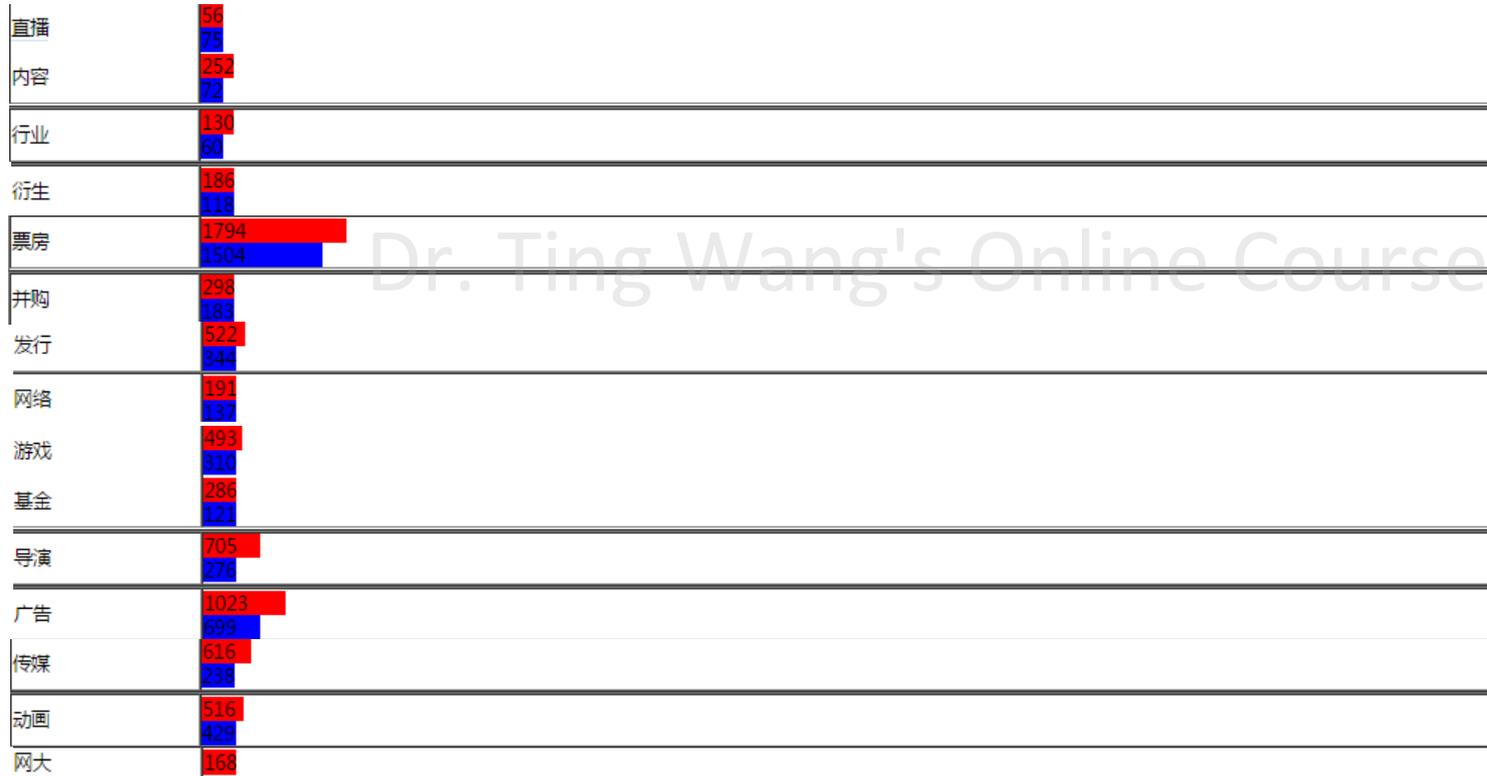


媒体热度



Film Box Office Prediction

Keyword Comparison (Industries)



H. BROTHERS
華 誼 兄 弟



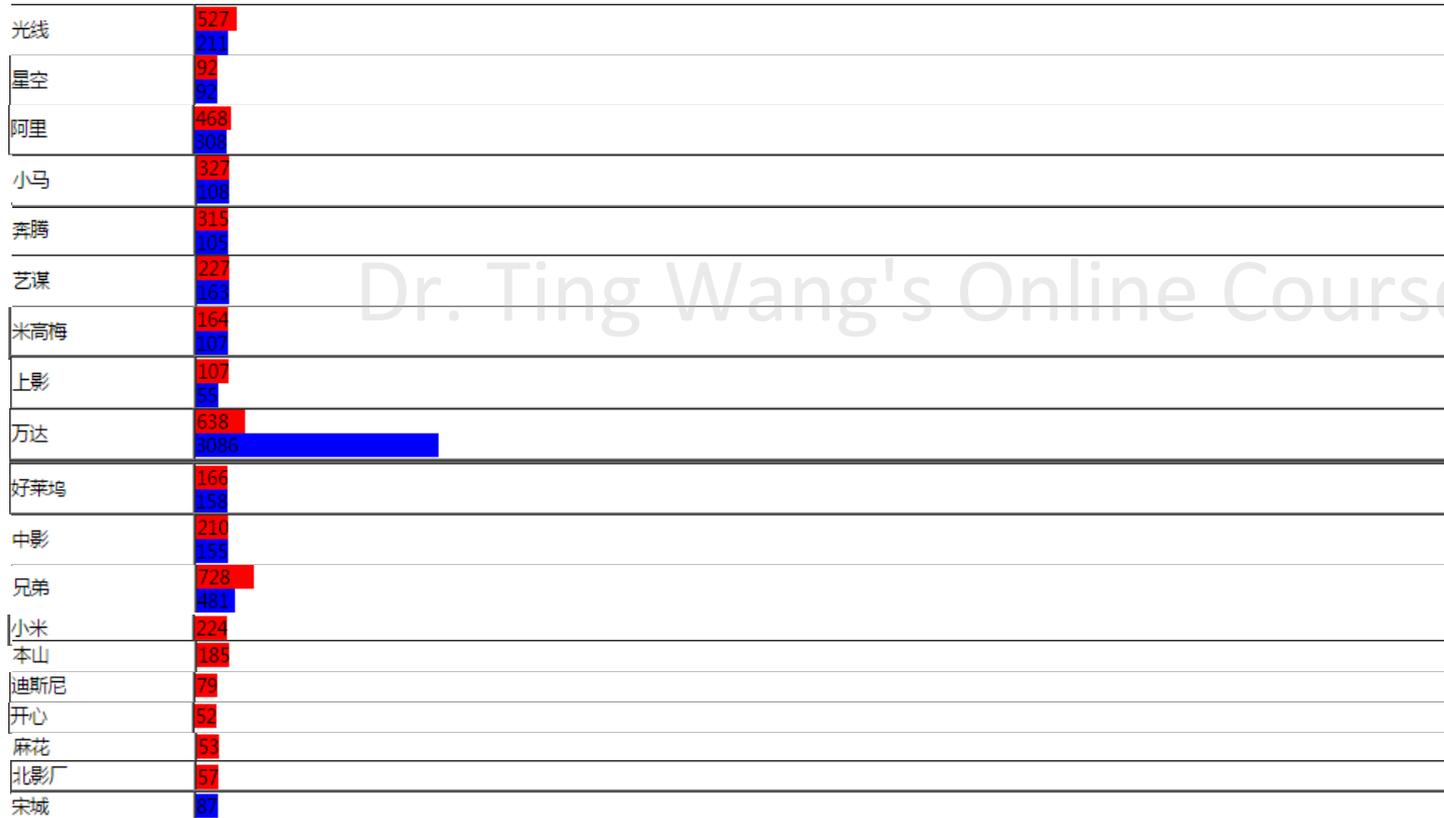
WANDA PICTURES



Dr. Ting Wang's Online Courses

Film Box Office Prediction

Keyword Comparison (Business Cooperation)



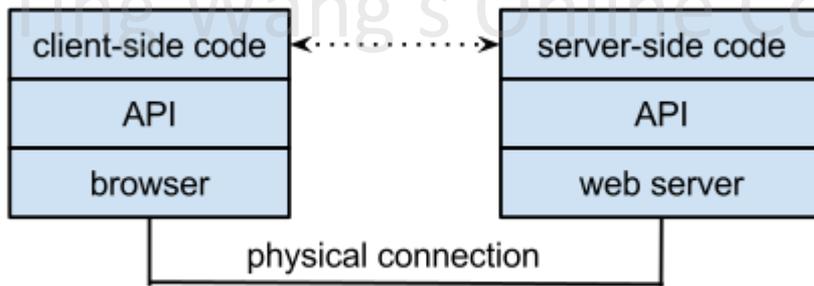
Dr. Ting Wang's Online Courses

Data Collection with API

API (Application Programming Interface)

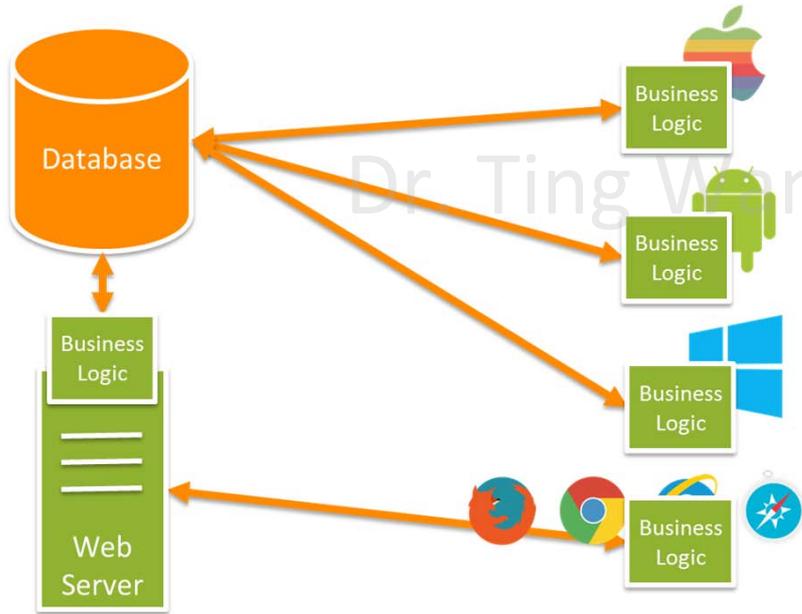
a set of subroutine definitions, protocols, and tools

What is Web API?

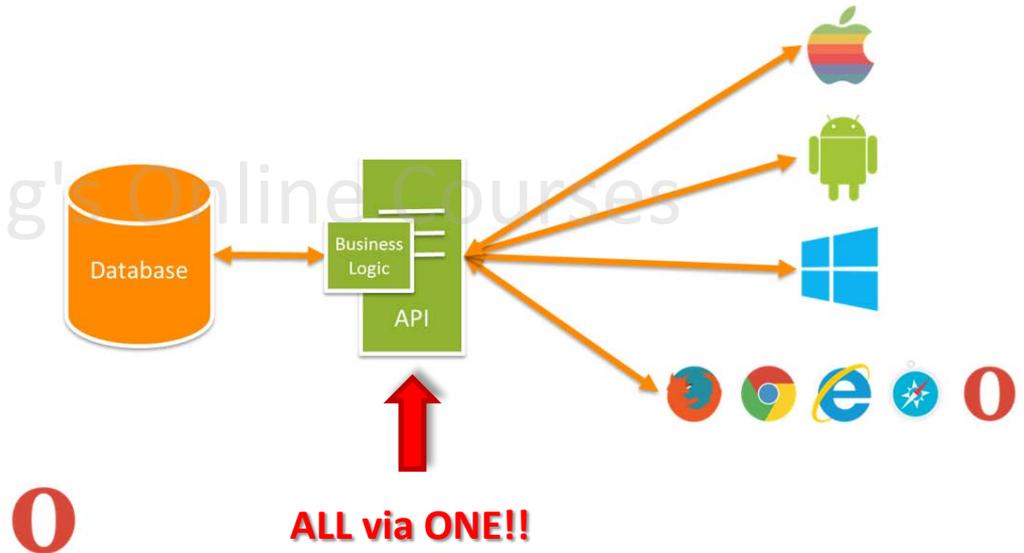


Data Collection with API

Short Short Ago



Now

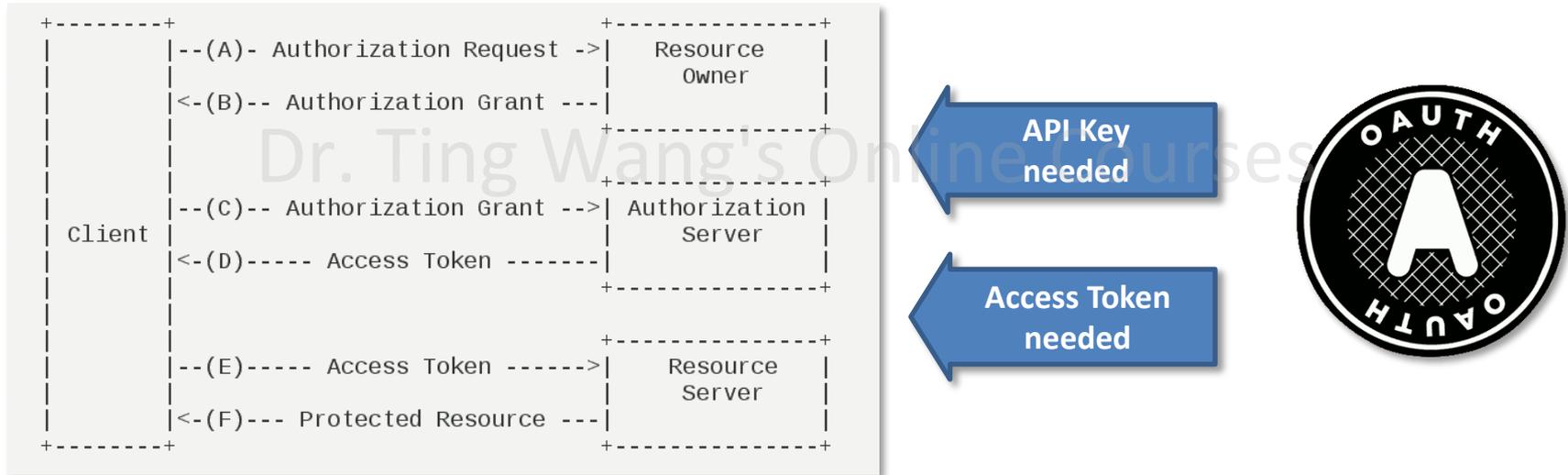


Integrating all the operations into one package, including "insert", "delete", "update", and "select".

Data Collection with API

OAuth

An open protocol to allow secure API authorization in a simple and standard method from web, mobile and desktop applications.



Ref. http://www.ruanyifeng.com/blog/2014/05/oauth_2_0.html

Data Collection with API

EXAMPLE 7:
Friendship of TVs on Twitter

I ♥ APIS

Get this free sticker at opigee.com

Friendship of TVs on Twitter

Problem Description

This research aims to investigate the significance and activeness of TV media companies in the world. Twitter friendships between TV broadcasting companies were employed for this research. Totally, about 104 TV channels from different countries are selected for this research. In-Degree and Out-Degree are used as features in this study.

Dr. Ting Wang's Online Courses

Features:

In-Degree: Followed by Others, which shows the significance to the world

Out-Degree: Following Others, which shows the activeness in the world

Friendship of TVs on Twitter

twitter and Its API

twitter

Do NOT tell me you do NOT know twitter, even you are a Chinese.

If you really do not know it, visit <https://twitter.com> right now!

Dr. Ting Wang's Online Courses



twitter API: <https://dev.twitter.com/>



Friendship of TVs on Twitter

Tweepy

An easy-to-use Python library for accessing the Twitter API.

Installation

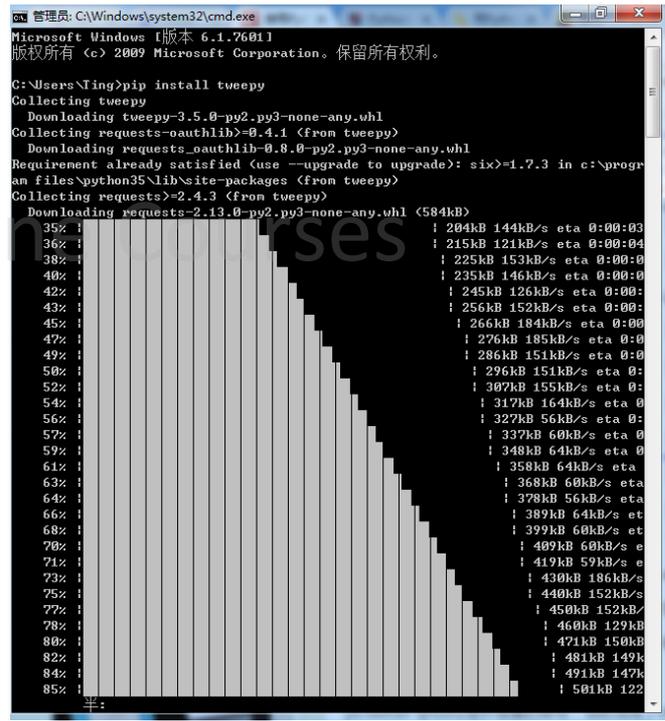
```
pip install tweepy
```

Reference to Tweepy

<https://github.com/tweepy/tweepy>

<http://www.tweepy.org/>

<http://pythonhosted.org/tweepy/>



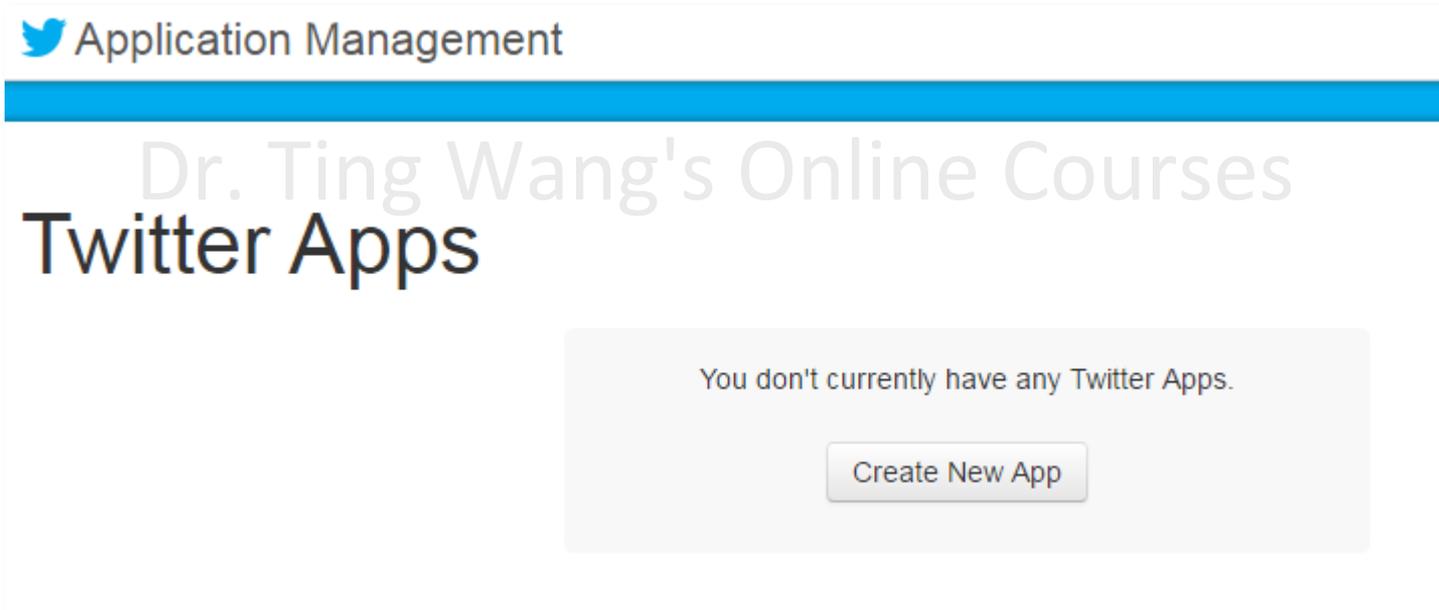
```
管理员: C:\Windows\system32\cmd.exe
Microsoft Windows [版本 6.1.7601]
版权所有 (c) 2009 Microsoft Corporation. 保留所有权利。

C:\Users\Ting>pip install tweepy
Collecting tweepy
  Downloading tweepy-3.5.0-py2.py3-none-any.whl
Collecting requests-oauthlib<=0.4.1 (from tweepy)
  Downloading requests_oauthlib-0.8.0-py2.py3-none-any.whl
Requirement already satisfied (use --upgrade to upgrade): six>=1.7.3 in c:\program files\python35\lib\site-packages (from tweepy)
Collecting requests>=2.4.3 (from tweepy)
  Downloading requests-2.13.0-py2.py3-none-any.whl (584kB)
35% |#####| 204kB 144kB/s eta 0:00:03
36% |#####| 215kB 121kB/s eta 0:00:04
38% |#####| 225kB 153kB/s eta 0:00:04
40% |#####| 235kB 146kB/s eta 0:00:04
42% |#####| 245kB 126kB/s eta 0:00:04
43% |#####| 256kB 152kB/s eta 0:00:04
45% |#####| 266kB 184kB/s eta 0:00:04
47% |#####| 276kB 185kB/s eta 0:00:04
49% |#####| 286kB 151kB/s eta 0:00:04
50% |#####| 296kB 151kB/s eta 0:00:04
52% |#####| 307kB 155kB/s eta 0:00:04
54% |#####| 317kB 164kB/s eta 0:00:04
56% |#####| 327kB 56kB/s eta 0:00:04
57% |#####| 337kB 60kB/s eta 0:00:04
59% |#####| 348kB 64kB/s eta 0:00:04
61% |#####| 358kB 64kB/s eta 0:00:04
63% |#####| 368kB 60kB/s eta 0:00:04
64% |#####| 378kB 56kB/s eta 0:00:04
66% |#####| 389kB 64kB/s eta 0:00:04
68% |#####| 399kB 60kB/s eta 0:00:04
70% |#####| 409kB 60kB/s eta 0:00:04
71% |#####| 419kB 59kB/s eta 0:00:04
73% |#####| 430kB 186kB/s eta 0:00:04
75% |#####| 440kB 152kB/s eta 0:00:04
77% |#####| 450kB 152kB/s eta 0:00:04
78% |#####| 460kB 129kB/s eta 0:00:04
80% |#####| 471kB 150kB/s eta 0:00:04
82% |#####| 481kB 149kB/s eta 0:00:04
84% |#####| 491kB 147kB/s eta 0:00:04
85% |#####| 501kB 122kB/s eta 0:00:04
```

Friendship of TVs on Twitter

Register Your Twitter APPs

Step 1 : visit <https://apps.twitter.com/>, and click on the button “Create New APP”.



Friendship of TVs on Twitter

Step2 : Input the detailed information of your new app.

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.

(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Friendship of TVs on Twitter

Step3: Get your API keys and Access Tokens

[Notes]

You may get all these information when you click on “**Keys and Access Tokens**”.



A screenshot of the Twitter developer console 'Application Settings' page for the user 'tingandwang'. The page shows the application's configuration, including access level, consumer key, and various URLs.

Organization
Information about the organization or company associated with your application

Organization	None
Organization website	None

Application Settings
Your application's Consumer Key and Secret are used to authenticate

Access level	Read and write (modify app permissions)
Consumer Key (API Key)	qcQYzdOBWYMArsRHp3vTbogC tokens
Callback URL	None
Callback URL Locked	No
Sign in with Twitter	Yes
App-only authentication	https://api.twitter.com/oauth2/token
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize

Friendship of TVs on Twitter

Steps for the System

1. Connect to twitter by API Keys and Access Tokens
2. Find the relationship using Tweepy functions

```
API.show_friendship(source_id/source_screen_name, target_id/target_screen_name)
```

Returns detailed information about the relationship between two users.

- Parameters:
- `source_id` - The user_id of the subject user.
 - `source_screen_name` - The screen_name of the subject user.
 - `target_id` - The user_id of the target user.
 - `target_screen_name` - The screen_name of the target user.

Return type: `Friendship` object

3. Save the results

Friendship of TVs on Twitter

Results

In-Degree for the significance of TV media

In-Degree Ranking

	Country	TV Media	Twitter ID	In-Degree
1	USA	CNN	CNN	26
2	Russia	Russia Today	RT_com	14
3	UK	BBC	BBCNews	13
4	Qatar	Al Jazeera television	AJENews	6
5	Spain	Radio Televisión Española	rtve	5
6	Netherland	Nederlandse Omroep Stichting	NOS	2
7	South Africa	South African Broadcasting Corporation, SABC	SABCNewsOnline	2
8	Lebanon	Al-Manar TV	almanarnews	2

Friendship of TVs on Twitter

Out-Degree for the activeness of TV media

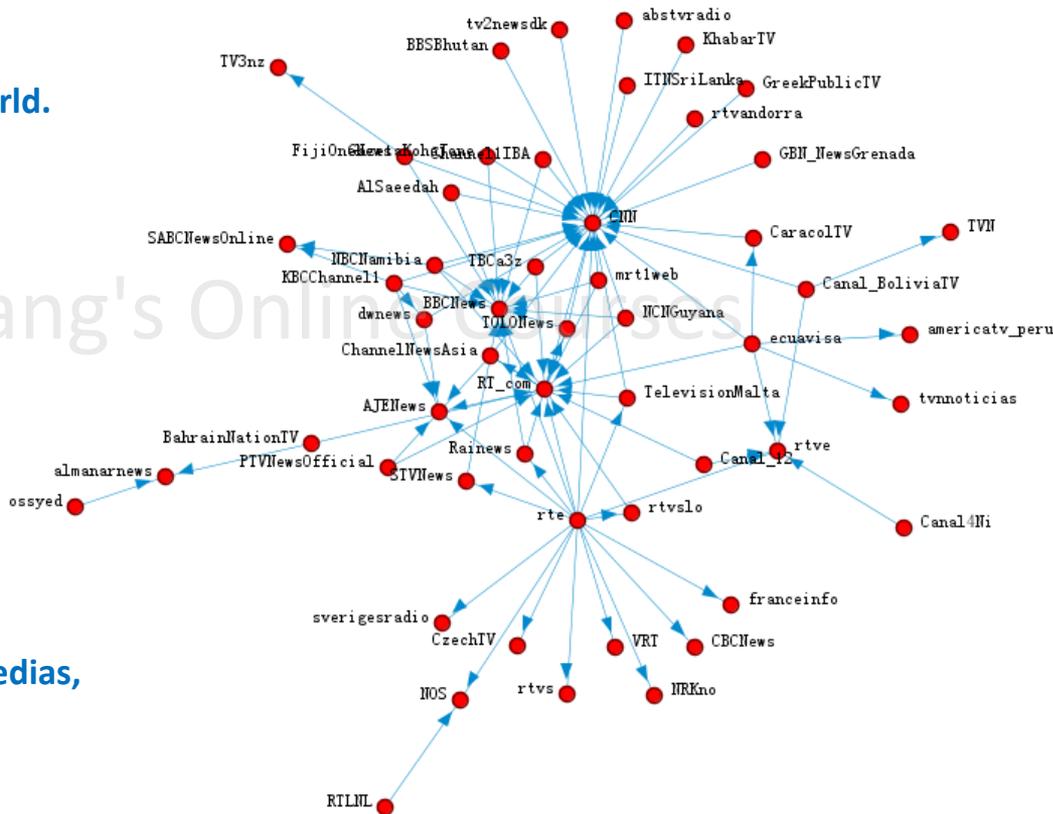
Out-Degree Ranking

	Country	TV Media	Twitter ID	Out-Degree
1	Ireland	Raidió Teilifís Éireann, Ireland's National Public Service Multi-Media Organisation	rte	17
2	Ecuador	Ecuavisa	ecuavisa	6
3	Kenya	Kenya Broadcasting Corporation	KBCChannel1	5
4	Namibia	Namibian Broadcasting Corporation	NBCNamibia	4
5	Russia	Russia Today	RT_com	4
6	Bolivia	Bolivia TV	Canal_BoliviaTV	3
7	Singapore	Television Corporation of Singapore	ChannelNewsAsia	3
8	Fiji	Fiji Television Limited	FijiOneNews	3
9	Guyana	National Communications Network, Guyana	NCNGuyana	3
10	Tonga	Tonga Broadcasting Commission	TBCa3z	3
11	Afghanistan	TOLONews	TOLONews	3
12	Macedonia	Macedonian Radio-Television	mrt1web	3

Friendship of TVs on Twitter

Conclusions

1. CNN, BBCNews, and Russia Today have the greatest significance to the world.
2. Raidió Teilifís Éireann from Ireland is the most active TV media.
3. Media from developed countries have greater influence.
4. Media from developing countries are more active.
5. CCTV, China neither follow any other medias, nor be followed by any other medias.





The End

Ting Wang's Online Courses

Thank You



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