前置技能：python基本语法，python环境搭建，requsets库安装与使用，正则表达式基本语法

源代码:（关键处都已经打了注释）

import requests

import re

import urllib.request

def getHTMLText(url): #获取网页文本

try:

r = requests.get(url, timeout=10)

r.raise\_for\_status()

r.encoding = r.apparent\_encoding

return r.text

except:

return ""

def parsePage(n,ilt, html): #下载网页上的图片

try:

plt = re.findall(r'g-search.\*?jpg',html) #正则表达式（匹配：以g-search开头，以jpg结尾的字符串）

for i in range(len(plt)):

try:

fo=open(str(n)+"."+str(i)+".jpg","wb")

r = requests.get('http://'+str(plt[i]), timeout=5);

iturl="http://"+str(plt[i])

web = urllib.request.urlopen(iturl)

itdata = web.read()

fo.write(itdata)

print("完成一次下载")

except:

print("错误")

except:

print("")

def main():

goods = '汉服' #指定淘宝上的一种商品名称，可更改为书，笔等任意词语

depth = 3 #递归深度，越大下载的图片越多

start\_url = 'https://s.taobao.com/search?q=' + goods

infoList = []

for i in range(depth):

try:

url = start\_url + '&s=' + str(44\*i)

html = getHTMLText(url)

parsePage(i,infoList, html)

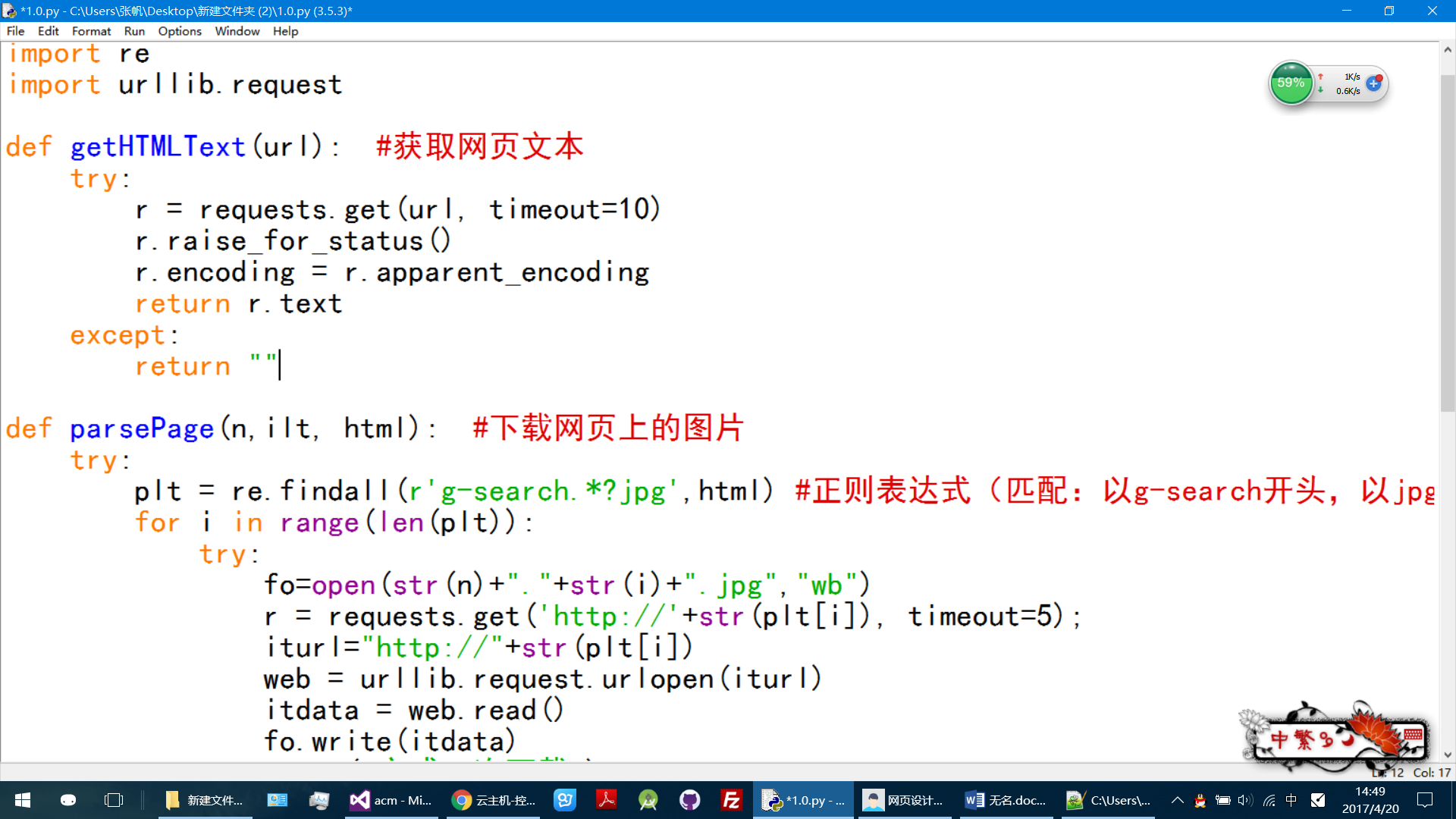
except:

continue

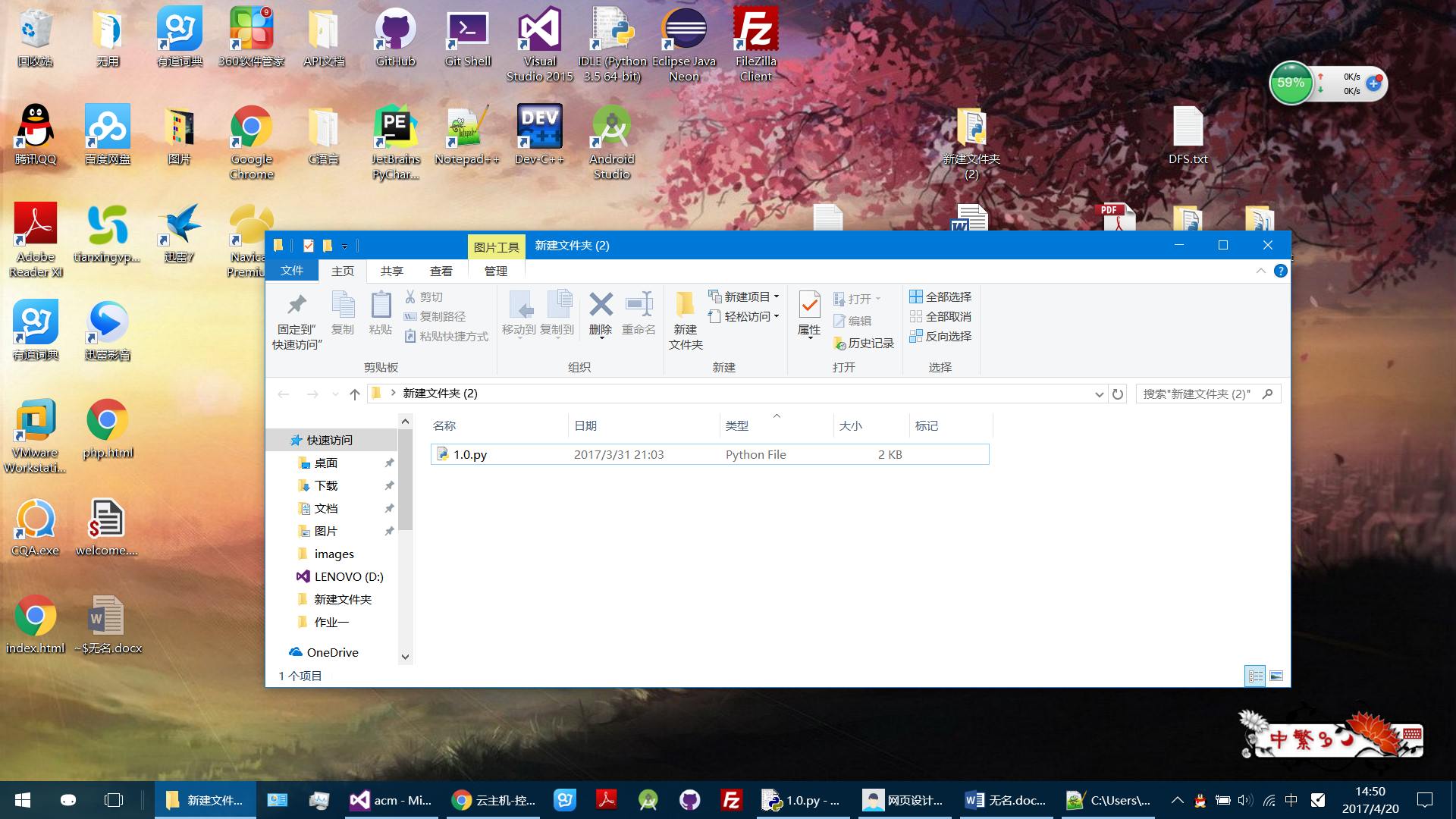
main()

图片：

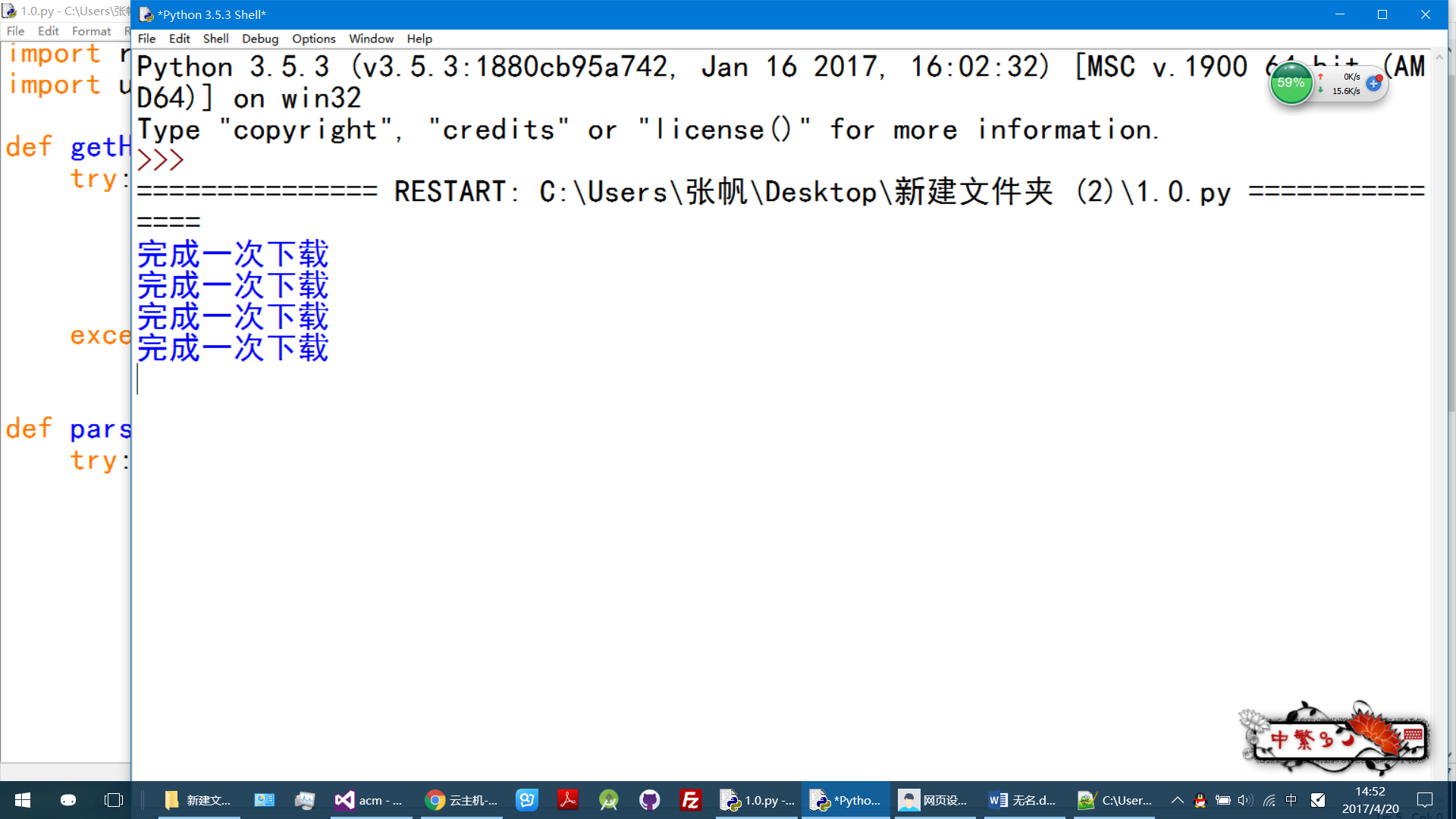
IDLE中代码：

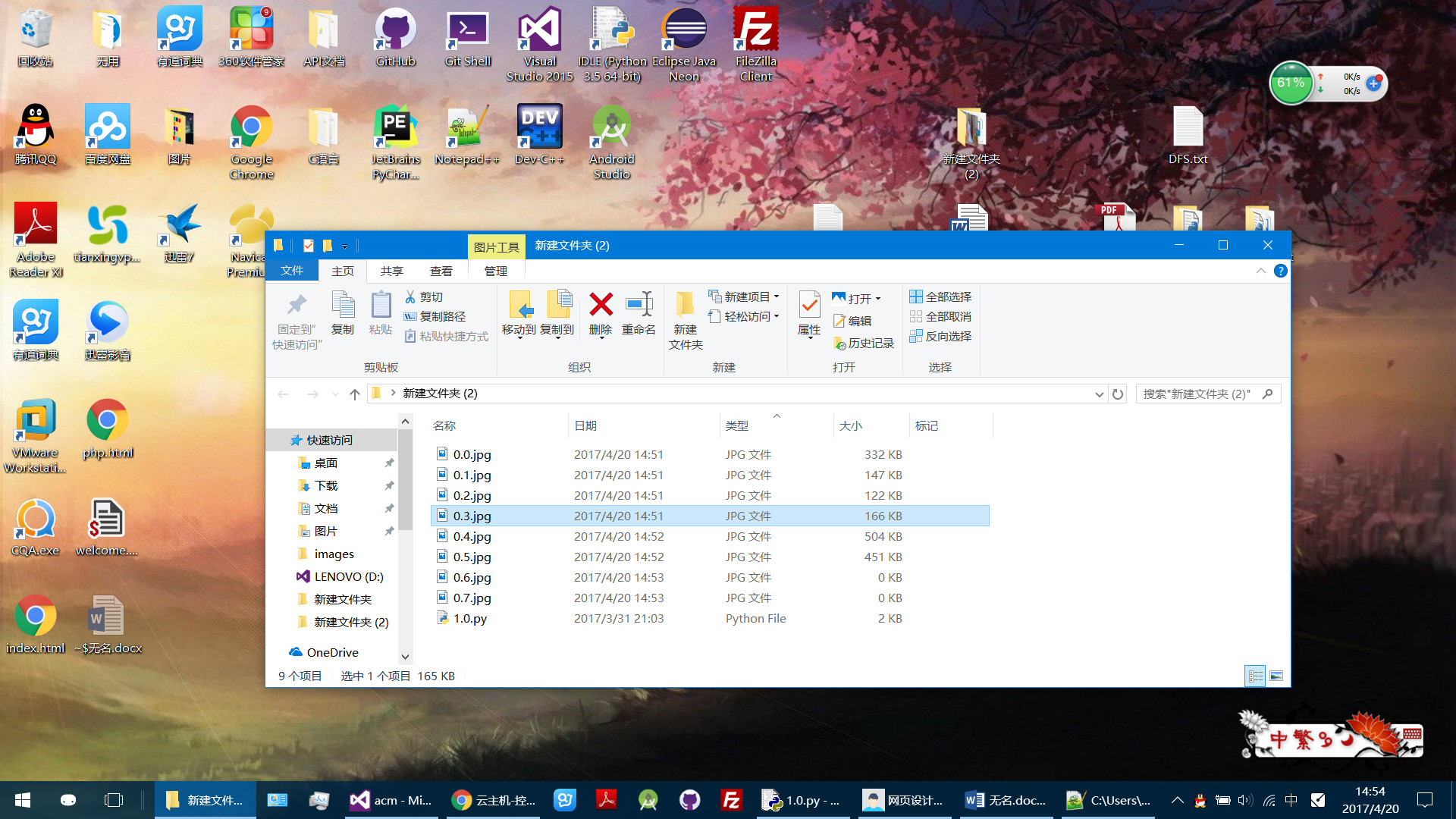


原文件夹：



运行中图片：

运行后文件夹：

下载的图片效果：

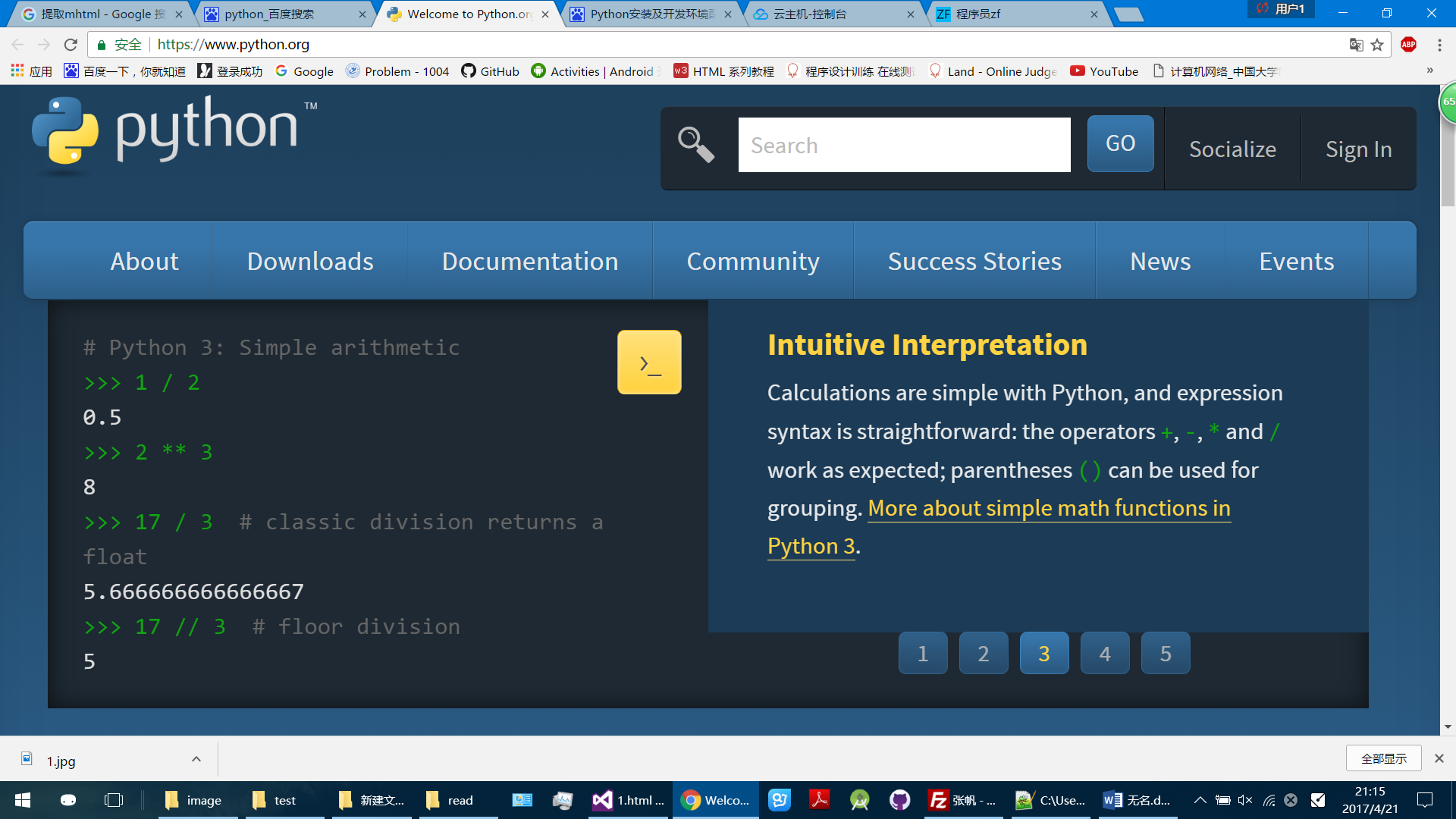




:

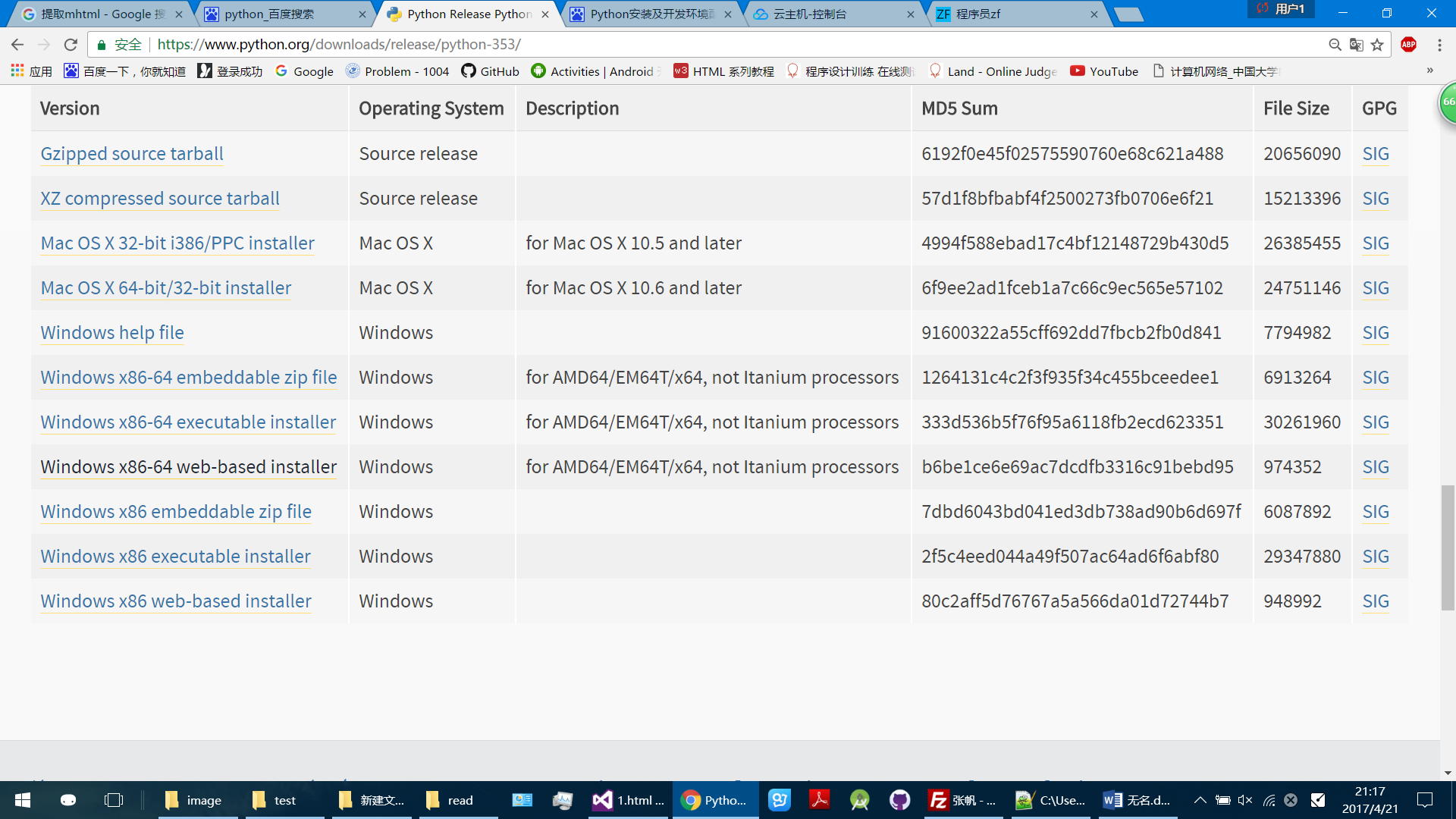
Python3.5环境搭建：

1. 进入python官网<https://www.python.org/>



1. 点击download

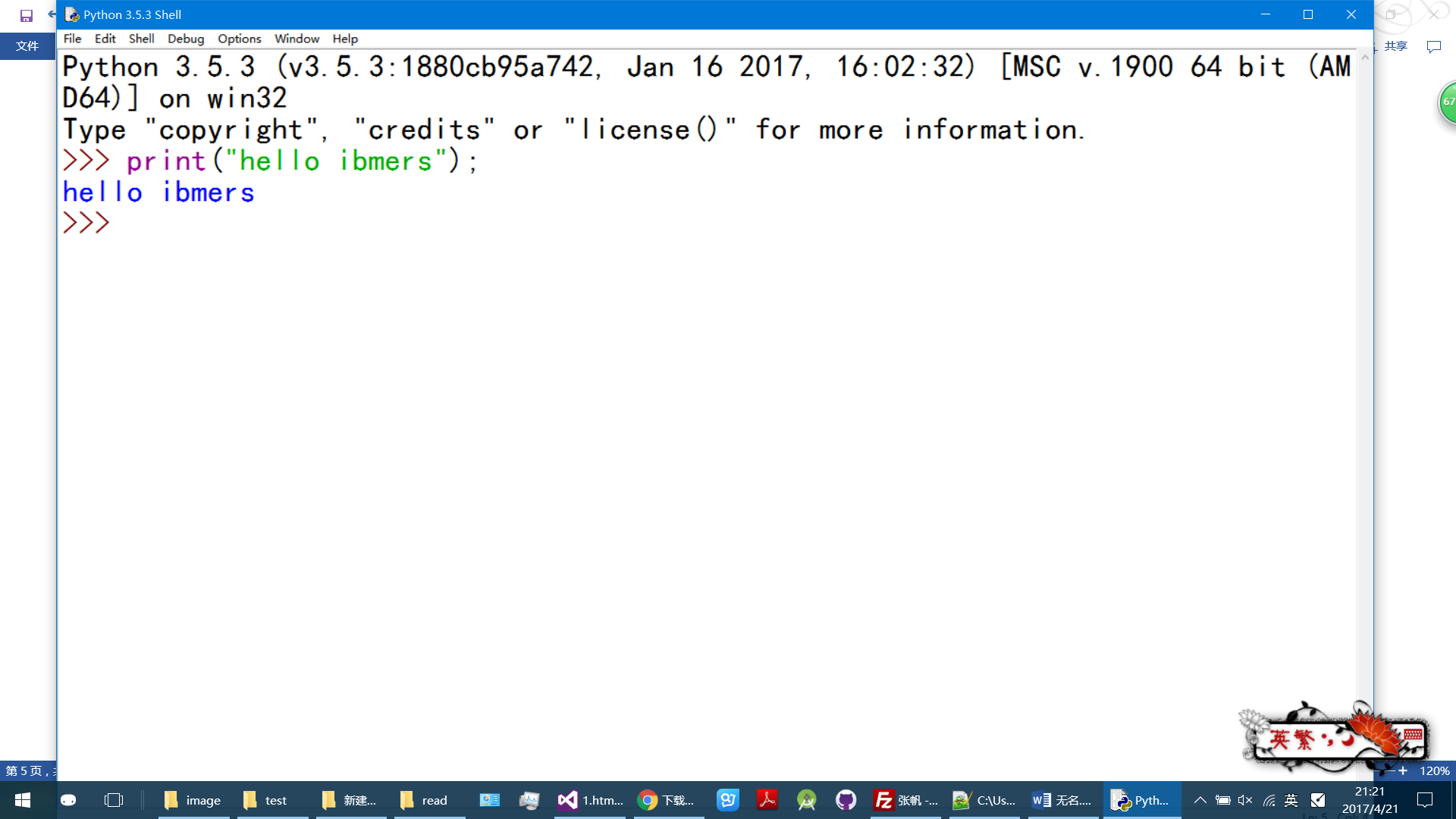
选择对应的版本下载。



然后一路安装下去就行了（我已经安装了，所以不好截图）

1. 通过python自带的IDLE就可以进行python程序的编写了

例如：



正则表达式基础：

正则表达式，又称正规表示式、正规表示法、正规表达式、规则表达式、常规表示法（英语：Regular Expression，在代码中常简写为regex、regexp或RE），计算机科学的一个概念。正则表达式使用单个字符串来描述、匹配一系列匹配某个句法规则的字符串。在很多文本编辑器里，正则表达式通常被用来检索、替换那些匹配某个模式的文本。（摘自维基百科）

正则表达式里的元字符及其作用：

| **元字符** | **说明** |
| --- | --- |
| . | 代表任意字符 |
| [ ] | 匹配内部的任一字符或子表达式 |
| [^] | 对字符集和取非 |
| - | 定义一个区间 |
| \ | 对下一字符取非（通常是普通变特殊，特殊变普通） |
| \* | 匹配前面的字符或者子表达式0次或多次 |
| \*? | 惰性匹配上一个 |
| + | 匹配前一个字符或子表达式一次或多次 |
| +? | 惰性匹配上一个 |
| ? | 匹配前一个字符或子表达式0次或1次重复 |
| {n} | 匹配前一个字符或子表达式 |
| {m,n} | 匹配前一个字符或子表达式至少m次至多n次 |
| {n,} | 匹配前一个字符或者子表达式至少n次 |
| {n,}? | 前一个的惰性匹配 |
| ^ | 匹配字符串的开头 |
| \A | 匹配字符串开头 |
| $ | 匹配字符串结束 |
| [\b] | 退格字符 |
| \c | 匹配一个控制字符 |
| \d | 匹配任意数字 |
| \D | 匹配数字以外的字符 |
| \t | 匹配制表符 |
| \w | 匹配任意数字字母下划线 |
| \W | 不匹配数字字母下划线 |

Python 的正则表达式的模块是 ‘re’, 它的基本语法规则就是指定一个字符序列，比如你要在一个字符串 s=’123abc456’ 中查找字符串 ’abc’, 只要这样写：

>>> import re

>>> s='123abc456eabc789'

>>> re.findall(r’abc’,s)

结果就是：

['abc', 'abc']