**一、python开发环境搭建**

**1.1 安装python-Windows环境**

首先，从Python的官方网站www.python.org下载最新的2.7.6版本，地址是这个：  
http://www.python.org/ftp/python/2.7.6/python-2.7.6.msi

然后，运行下载的MSI安装包，不需要更改任何默认设置，直接一路点“Next”即可完成安装：

默认会安装到C:\Python27目录下，但是当你兴致勃勃地打开命令提示符窗口，敲入python后，会得到

‘python’不是内部或外部命令，也不是可运行的程序或批处理文件。

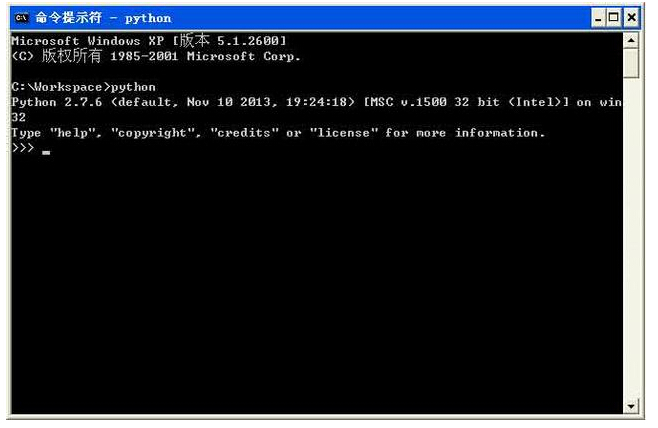
这是因为Windows会根据一个Path的环境变量设定的路径去查找python.exe，如果没找到，就会报错。解决办法是把python.exe所在的路径C:\Python27添加到Path中。

在控制面板中打开“系统属性”，点击“高级”，“环境变量”，打开“环境变量”窗口，在系统变量中，找到“Path”变量，然后点击“编辑”：



在“编辑系统变量”的窗口中，可以看到，变量名是Path，在变量值的最后面，先添加一个分号“;”（注意用英文输入法，千万不要输入中文分号），再写上C:\Python27（如果安装的时候没有更改过安装目录），然后连续点“确定”，“确定”，“确定”把所有窗口都关掉。

现在，再打开一个新的命令行窗口（**一定要关掉原来的命令行窗口，再新开一个**），输入python：



看到上面的画面，就说明Python安装成功！

你看到提示符>>>就表示我们已经在Python交互式环境中了，可以输入任何Python代码，回车后会立刻得到执行结果。现在，输入exit()并回车，就可以退出Python交互式环境（直接关掉命令行窗口也可以！）。

1.2 安装python依赖包

pip安装（PyCharm 2017.3.3）

pip的作用：安装python所需的包

安装pip位置：D:\Python36\Scripts

cmd 命令：easy\_install pip

pip安装本地资源？

pip安装包的源文件的存放路径？

通过pip install -r requirements.txt一键安装所有的依赖包

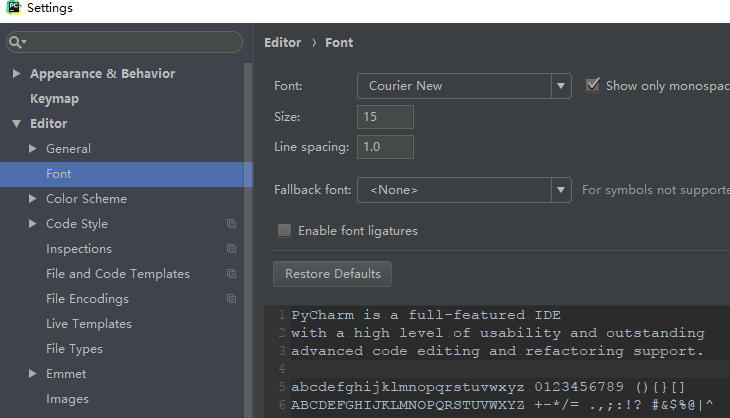
Flask：Python web 框架

requests:Python HTTP客户端库

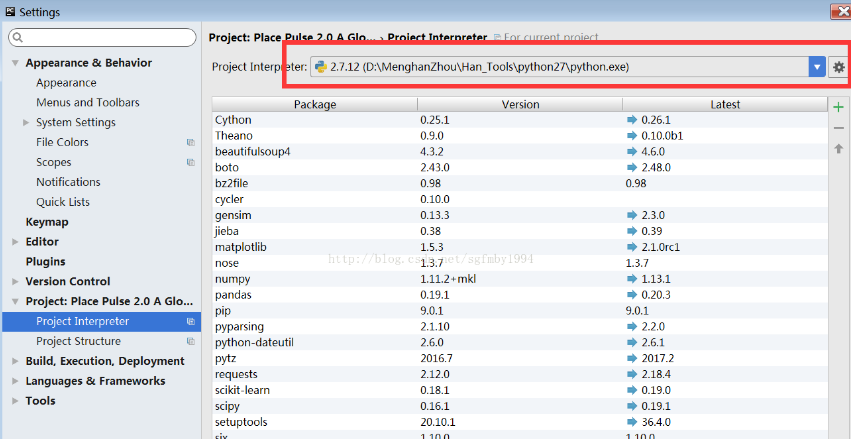
1.3 安装与配置python IDE——pycharm

设置pycharm字体

File——》settings——》（快捷：ctrl+alt+s）



在pycharm中切换python版本的方法



Pycharm快捷键

序号 快捷键 作用

1 ALT + SHIFT + F10 选择程序文件并运行代码

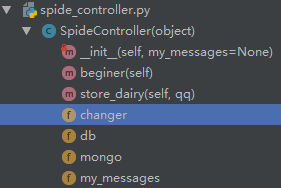
2 ALT + SHIFT + F9 选择程序文件并调试代码

3 SHIFT + F10 运行代码

4 SHIFT + F9 调试代码

5 CTRL + SHIFT + F10 运行当前编辑区的程序文件

Pycharm代码元素代号



p：parameter 参数

m：method 方法

c：class 类

v：variable 变量

f：function 函数

**python的常用领域demo分析**

x.1金融证券

****

x.2 网页爬虫

demo1 58同城爬虫

功能：

# 方法1：获取产品分类列表

# 方法2：获取单个分类的产品列表

# 方法3：获取单个产品的详细信息

技术栈：python2.7

主要使用 requests 模块下载页面，部分使用 BeautifulSoup对页面内容解析。

数据库中间件：pymongo

数据库：mongodb

|  |
| --- |
| requests  requests.get(start\_url)🡪  *return* request('get', *url*, params=*params*, \*\**kwargs*)🡪  *return* session.request(method=*method*, url=*url*, \*\**kwargs*)🡪  *return* self.send(prep, \*\*send\_kwargs)🡪  hooks = *hooks or* dict() hooks = *hooks*.get(*key*)  *return* dispatch\_hook('response', hooks, r, \*\**kwargs*)  （hooks:python事件响应钩子 event hooks） |

demo2 多线程qq空间爬虫

功能：对某个qq空间的访客、说说、日志、个人四种信息进行爬取，后期可以通过Gephi生成社交热力图

技术栈：python2.7

主要使用 requests 模块抓取，部分使用 BeautifulSoup 解析。

同一节点下的多个同级访客的多线程爬取使用 multiprocessing.dummy 。

登录qq空间需要抓取 Cookie， 使用 selenium 和 PhantomJS 。

判重使用 BitVector 。

数据库中间件：pymongo

数据库：mongodb

{

“代码结构”

Controller

spide\_controller

Service:”爬虫功能单元（下载页面、解析页面等方法）”

blog\_spider

friend\_spider

information\_spider

mood\_spider

Util

public\_methods

Config

init\_messages

myQQ.txt:”存放爬虫初始爬取的源QQ”

log

ghostdriver.log

QQFailSpided.txt:”爬取失败”

QQForSpider.txt:”待爬”

QQHadSpided.txt:”已爬”

Plugin

Phantomjs.exe:” 抓取 Cookie”

}

代码运行流程

1.设置程序统一编码格式

2. # 读取本地的爬虫信息，并对爬虫进行参数初始化

3. # spider对象控制生成多线程爬虫

4. # 运行爬虫

启动警告

UserWarning: Selenium support for PhantomJS has been deprecated, please use headless versions of Chrome or Firefox instead

x.3 人脸识别