## 一 创建scrapy项目与爬虫文件

### 创建scrapy项目

1. cd 到scrapy的父文件夹下
2. scrapy startproject project1

例如：

|  |
| --- |
|  |

### 创建spider文件

第一步：cmd进入项目文件夹

|  |
| --- |
|  |

第二步：新建项目

|  |
| --- |
|  |

第三步：编写爬虫文件

|  |
| --- |
|  |

## 二 入门案例

第一步：写spider文件

|  |
| --- |
| *# -\*- coding: utf-8 -\*-* import scrapy  class Myspider1Spider(scrapy.Spider):  name = 'myspider1'  allowed\_domains = ['taobao.com']  start\_urls = ['http://taobao.com/']   def parse(self, response):  print(response.body) |

第二步：settings中的去掉robot协议

|  |
| --- |
| * ROBOTSTXT\_OBEY = False |

第三步：cmd中运行spider文件

|  |
| --- |
|  |

## 三 在pycharm中执行scrapy

第一步：在spiders文件下新建一个start.py（名字随意）

|  |
| --- |
| import scrapy.cmdline as cmd  proj = input(“输入spider的名字”)  command = ‘scrapy crawl ’+proj  cmd.execute(command.split()) |

## 三 用scrapy爬百度贴吧

## 四 scrapy 框架流程

|  |
| --- |
|  |