CONCURRENT\_ITEMS

默认: 100

Item Processor(即 Item Pipeline) 同时处理(每个response的)item的最大值。

##### CONCURRENT\_REQUESTS

默认: 16

Scrapy downloader 并发请求(concurrent requests)的最大值。

##### DEFAULT\_REQUEST\_HEADERS

默认: 如下

{

'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,\*/\*;q=0.8',

'Accept-Language': 'en',

}

Scrapy HTTP Request使用的默认header。

DEPTH\_LIMIT

默认: 0

爬取网站最大允许的深度(depth)值。如果为0，则没有限制。

##### DOWNLOAD\_DELAY

默认: 0

下载器在下载同一个网站下一个页面前需要等待的时间。该选项可以用来限制爬取速度， 减轻服务器压力。同时也支持小数:

DOWNLOAD\_DELAY = 0.25 # 250 ms of delay

默认情况下，Scrapy在两个请求间不等待一个固定的值， 而是使用0.5到1.5之间的一个随机值 \* DOWNLOAD\_DELAY 的结果作为等待间隔。

DOWNLOAD\_TIMEOUT

默认: 180

下载器超时时间(单位: 秒)。

##### ITEM\_PIPELINES

默认: {}

保存项目中启用的pipeline及其顺序的字典。该字典默认为空，值(value)任意，不过值(value)习惯设置在0-1000范围内，值越小优先级越高。

ITEM\_PIPELINES = {

'mySpider.pipelines.SomethingPipeline': 300,

'mySpider.pipelines.ItcastJsonPipeline': 800,

}

LOG\_ENABLED

默认: True

是否启用logging。

LOG\_ENCODING

默认: 'utf-8'

logging使用的编码。

LOG\_LEVEL

默认: 'DEBUG'

log的最低级别。可选的级别有: CRITICAL、 ERROR、WARNING、INFO、DEBUG 。

##### USER\_AGENT

默认: "Scrapy/VERSION (+<http://scrapy.org>)"

爬取的默认User-Agent，除非被覆盖。

##### PROXIES： 代理设置

示例：

PROXIES = [

{'ip\_port': '111.11.228.75:80', 'password': ''},

{'ip\_port': '120.198.243.22:80', 'password': ''},

{'ip\_port': '111.8.60.9:8123', 'password': ''},

{'ip\_port': '101.71.27.120:80', 'password': ''},

{'ip\_port': '122.96.59.104:80', 'password': ''},

{'ip\_port': '122.224.249.122:8088', 'password':''},

]

##### COOKIES\_ENABLED = False

禁用Cookies