

# 一. 使用方法

- Step 1: `cd` 至 `.\NSD-jd-tm\jd\jdWare` 目录下;
- Step 2: `scrapy crawl jdSpider`。

# 二. 环境要求

- python 3;
- 安装scrapy;
- 安装mongodb, 且mongod服务打开。

# 三. 字段说明

**NSD\_JD** 数据库下有三个collections:

**SPU** collection:

- **SPU\_NUM**: 产品编号 (字符串);
- **SPU\_ID**: 产品ID (字符串);
- **SPU\_TIT**: 产品标题 (字符串);
- **SPU\_DESP**: 产品描述 (数组);
- **SPU\_SPEC**: 产品规格 (数组);
- **SPU\_RATE**: 产品口碑得分 (浮点数);
- **SPU\_REVNUM**: 页面显示总评论数 (整数);
- **SPU\_TREVNUM**: 排除默认评论后的实际评论数 (整数);

**SKU** collection:

- **SPU\_NUM**: 产品编号 (字符串);
- **SKU\_ID**: 版本ID (字符串);
- **SKU\_PRICEBF**: 版本原价 (浮点数);
- **SKU\_PRICEAF**: 版本现价 (浮点数);
- **SKU\_TIT**: 版本分类 (字符串);

**REVIEW** collection:

- **SPU\_NUM**: 产品编号 (字符串);

- `SPU_TIT`：产品标题（字符串）；
- `U_NAME`：用户名称（字符串）；
- `U_LEVEL`：用户等级（字符串）
- `SKU_TIT`：购买的产品类型（字符串）；
- `REV_TIME`：评论时间（字符串）；
- `REV_CONTENT`：评论内容（字符串）；
- `REV_SCORE`：购买评价分数（整数）；
- `REV_PIC`：是否晒图（1为有图）；

## 四. 其他

- 每个关键字搜索，最多搜索到75\*100，即7500个结果；
- 评论部分，由于存在后期访问不稳定的原因，爬取到的评论并非全部评论，目前是按“多多益善”的原则，尽可能爬取多的评论。