唯品会json爬虫

1 UPDATE 21-1-14-17:21

总爬虫为cralwer.py

总输出为开头为 final-merged-data-{关键词}.json

2 注意事项

- 唯品会所有数据都记录在json格式里面,包括各种评论信息以及图片URL,因此不需要下载html
- 如有必要,请自行学习json格式,不难
- 需要安装 `playwright`,请自行安装

3 食用方法

运行cralwer.py,输入关键词和页数(都有提示)

喝一杯茶(速度大约为5分钟-120个商品)

```
 1  "6919592689722022353": {
 2          "spuId": "3482766938583883811",
 3          "productId": "6919592689722022353",
 4          "brandId": "1710612817",
 5          "title": "新款男子NIKE AIR MAX 2090 C/S经典运动休闲鞋",
 6          "vipshopPrice": "680",
 7          "marketPrice": "999",
 8          "smallImage":
    "http://h2.appsimg.com/a.appsimg.com/upload/merchandise/pdcvis/2021/09/26/181/137b9a6e-
    1a49-4a24-9567-30a7acfd58d4_420_531.jpg",
 9          "brandStoreName": "Nike",
10          "brandStoreLogo":
    "http://a.vpimg3.com/upload/brandcool/0/3ccd836975e4471992b1dd58c25145bf/10000630/prima
    ry.png",
11          "brandStoreWhiteLogo":
    "http://h2.appsimg.com/a.appsimg.com/upload/brandcool/0/13813886e33849ec9a9087270261235
    e/10000630/white.png",
12          "images": [
13              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/150/c616b07b-
    2959-4c30-971d-bf06078d3be4.jpg",
14              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/13/aa1ff85d-436d-
    4d7d-ab2a-8b525fa52daf.jpg",
15              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/89/7ac4bdc7-1481-
    43de-b116-372f7710f4c3.jpg",
16              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/19/6ba63b8d-ff2a-
    40bd-972e-94c9bc5ecbcb.jpg",
17              ...
18              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/94/0dfa8454-b45d-
    480f-b007-c0e80cccec9c.jpg",
```

```
19              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/40/1135984b-cad4-
        4fa8-8b97-defbbbc8da2e.jpg",
20              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/61/7216397e-6a33-
        4c05-b7d1-58ea5252fb47.jpg",
21              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/131/de76e240-
        abfe-4541-beb0-85893361ed70.jpg",
22              "http://a.vpimg4.com/upload/merchandise/pdcvis/2021/09/26/187/2ec8a862-
        a63d-48d7-a2a0-5ba30953db56.jpg"
23          ],
24          "comments": [//这里只有两条不大有价值的评论
25              {
26                  "content": "合适",
27                  "nlpScore": 50,
28                  "satisfiedStatus": "VERY_SATISFIED"
29              },
30              {
31                  "content": "39",
32                  "nlpScore": 50,
33                  "satisfiedStatus": "VERY_SATISFIED"
34              }
35          ],
36          "brandStoreSn": "10000630",
37          "brandShowName": "Nike",
38          "squareImage":
        "http://h2.appsimg.com/a.appsimg.com/upload/merchandise/pdcvis/2021/09/26/94/0dfa8454-
        b45d-480f-b007-c0e80cccec9c.jpg",
39          "logo":
        "http://a.vpimg3.com/upload/brandcool/0/3ccd836975e4471992b1dd58c25145bf/10000630/prima
        ry.png",
40          "attrs": [//各种属性
41              {
42                  "name": "类型",
43                  "value": "休闲鞋"
44              },
45              {
46                  "name": "跟型",
47                  "value": "平底"
48              },
49              {
50                  "name": "里绒情况",
51                  "value": "无"
52              },
53              {
54                  "name": "鞋底软度",
55                  "value": "适中"
56              },
57              {
58                  "name": "闭合方式",
```

```
                    "value": "系带"
            },
            {
                    "name": "风格",
                    "value": "运动风"
            },
            {
                    "name": "鞋帮",
                    "value": "低帮"
            },
            {
                    "name": "适用季节",
                    "value": "春,秋,冬"
            },
            {
                    "name": "适用人群",
                    "value": "青少年,青年,中年"
            },
            {
                    "name": "选购热点",
                    "value": "主推款"
            },
            {
                    "name": "功能",
                    "value": "耐磨"
            },
            {
                    "name": "图案",
                    "value": "LOGO"
            },
            {
                    "name": "面材质",
                    "value": "织物"
            },
            {
                    "name": "适用项目",
                    "value": "日常"
            },
            {
                    "name": "鞋底材质",
                    "value": "橡胶底"
            },
            {
                    "name": "科技",
                    "value": "无"
            },
            {
                    "name": "适用性别",
```

```
            "value": "男士"
        },
        {
            "name": "详细材质信息",
            "value": "详见商品吊牌"
        }
    ]
},
```