



链滴

小小爬虫

作者: [momoda](#)

原文链接: <https://ld246.com/article/1519207109696>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

简单爬虫

代码地址 <https://github.com/zousiliang/pacon>

最近花了一点时间研究一下爬虫，没有使用第三方爬虫框架,当然你也可以使用第三方框架 如:

- 1.分布式爬虫: Nutch
- 2.JAVA单机爬虫: Crawler4j、WebMagic、WebCollector
- 3.非JAVA单机爬虫: scrapy、pyspider

废话不多说,先看结果,爬取熊猫tv主播数据 并进行降序排列:

rank 1 : Misaya若风lol : 89.7万
rank 2 : 贾克虎、虎神 : 62.7万
rank 3 : 蓝战非 : 60.2万
rank 4 : 瓦莉拉的小伙伴 : 57.8万
rank 5 : ToveLo : 32.5万
rank 6 : 君克解说 : 24.2万
rank 7 : 风行云 : 21.6万
rank 8 : RNG、MLXG : 20.9万
rank 9 : QG无敌老飞牛 : 14.8万
rank 10 : 爷大千、 : 13.4万
rank 11 : 萝莉图阿 : 12.2万
rank 12 : 一阵雨不是一阵奶 : 10.0万
rank 13 : LOL、摇摆哥 : 7.5万
rank 14 : 真正的村长ALEX : 6.8万
rank 15 : RNG1letme1 : 6.4万
rank 16 : lol稳贱骨炼金 : 6.0万

爬虫文件为:

spider/spider.py

执行代码为:

python3 spider.py