

# Contents

1 简介 (Introduction)	1
2 快速上手 (Getting Started)	1
2.1 环境准备	1
3 添加新的代理源	1

## 1 简介 (Introduction)

用于抓取代理，数据来自代理供应商的网站，通过解析这些网站获得的 `ip:port`。并用一个验证脚本不断验证所抓到的 IP，把能够使用的标记出来。

## 2 快速上手 (Getting Started)

### 2.1 环境准备

有两种配置环境的方式 \* 本机上配置

```
# 这一步自行根据包管理器的类型来安装
install nodejs 7.6+ & install yarn
```

```
# 安装依赖
yarn add sprintf-js & yarn add puppeteer & yarn add request
```

```
# 运行demo
node xicidaili.js
```

- 使用 docker 配置

```
# 拉取镜像
docker pull richardtt1225/proxypool
```

```
# 生成容器运行，/home/work/yzj/core/xicidaili.js 是本地文件路径，/app/index.js 是容器里面被执行文件
# 例子如下
```

```
docker run --rm -v /home/work/yzj/core/xicidaili.js:/app/index.js richardtt1225/proxypool
```

- 运行结果 demo:

```
[xicidaili]: 获得新的ip数: 75/100 [20180613065252]
```

## 3 添加新的代理源

目录下有个 `module.js`，这是个通用模板，复制一份模板，完成模板里面的三个 TODO 即可生成新的爬虫文件

- TODO1: 爬虫名字
- TODO2: 从网页上提取出 ip 块
- TODO3: 对 ip 块进行解析，分解成一个个 ip