

BaiduIndexHunter 安装与部署指南

!! 这个项目只是我个人在爬取百度指数数据进行实证的过程中，为了方便对众多关键词的批量爬取所构建的项目。本项目只能用于学习交流，不可用于商业用途，因商业用途导致的一系列后果自行承担。

一、环境要求

在开始配置之前,请确保你的电脑满足以下要求:

操作系统

- Windows 10/11
- macOS 10.14 或更高版本
- Linux (Ubuntu 18.04+ / CentOS 7+ / Debian 9+)

必需软件

1. Python 3.8 或更高版本
2. Node.js 14.0 或更高版本 (推荐 16.x 或 18.x)
3. MySQL 8.0 或更高版本
4. Redis 5.0 或更高版本

二、配置

1. 配置数据库

代码块

```
1  # 在mysql中运行baidu-index-hunter-backend\BaiduIndexHunter.sql语句创建数据库
2  1. cd BaiduIndexHunter\baidu-index-hunter-backend
3  2. mysql -u root -p BaiduIndexHunter < BaiduIndexHunter.sql
```

2. 配置环境变量

在baidu-index-hunter-backend目录下创建.env文件，然后将下面内容复制到.env中，并修改对应的参数

代码块

```
1  # =====
2  # MySQL 数据库配置
3  # =====
4  MYSQL_HOST=localhost
5  MYSQL_PORT=3306
6  MYSQL_USER=root
7  MYSQL_PASSWORD=你的MySQL密码
8  MYSQL_DB=BaiduIndexHunter
9
10 # =====
11 # Redis 配置
12 # =====
13 REDIS_HOST=localhost
14 REDIS_PORT=6379
15 REDIS_DB=0
16 REDIS_PASSWORD=
17
18 # =====
19 # Flask 应用配置
20 # =====
21 FLASK_HOST=0.0.0.0
22 FLASK_PORT=5001
23 FLASK_DEBUG=false
24
25 # =====
26 # API 配置
27 # =====
28 API_HOST=0.0.0.0
29 API_PORT=5001
30 API_DEBUG=True
31 API_SECRET_KEY=baidu_index_hunter_secret_key_请修改此密钥 #项目中暂时应该用不到，随
    便设置即可
32 API_TOKEN_EXPIRE=86400
33 API_CORS_ORIGINS=*
34
35 # =====
36 # 任务配置
37 # =====
38 MAX_CONCURRENT_TASKS=20
39 TASK_QUEUE_CHECK_INTERVAL=5
40 DEFAULT_TASK_PRIORITY=5
41 MAX_RETRY_COUNT=3
42 RETRY_DELAY=120
43
44 # =====
45 # 登录和健康检查配置
46 # =====
```

```
47 LOGIN_INTERVAL=86400
48 HEALTH_CHECK_INTERVAL=3600
49
50 # =====
51 # Cookie 检查配置
52 # =====
53 COOKIE_CHECK_INTERVAL=180
54 AB_SR_UPDATE_INTERVAL=3600
55 RESUME_TASK_CHECK_INTERVAL=300
56
57 # =====
58 # Cookie 管理配置
59 # =====
60 COOKIE_MIN_AVAILABLE_COUNT=10
61 COOKIE_MAX_LOGIN_RETRY=2
62 COOKIE_BLOCK_COOLDOWN=1800
63 COOKIE_EXPIRATION_BUFFER=3600
64 COOKIE_ROTATION_STRATEGY=round_robin
65 COOKIE_MAX_USAGE_PER_DAY=1800
66
67 # =====
68 # 爬虫配置
69 # =====
70 SPIDER_MIN_INTERVAL=0.1
71 SPIDER_MAX_INTERVAL=0.3
72 SPIDER_DEFAULT_INTERVAL=0.2
73 SPIDER_RETRY_TIMES=2
74 SPIDER_TIMEOUT=10
75 SPIDER_MAX_WORKERS=20
76 SPIDER_MAX_CONSECUTIVE_FAILURES=2
77 SPIDER_FAILURE_MULTIPLIER=1.1
78 SPIDER_USER_AGENT_ROTATION=True
79 SPIDER_PROXY_ENABLED=False
80 SPIDER_PROXY_URL=
81 SPIDER_PROXY_AUTH=
82
83 # =====
84 # 日志配置
85 # =====
86 LOG_LEVEL=INFO
87 LOG_RETENTION=7
88 LOG_FORMAT=%(asctime)s - %(levelname)s - %(name)s - %(message)s
89 LOG_FILE_SIZE=10485760
90 LOG_BACKUP_COUNT=5
91
92 # =====
93 # 输出配置
```

```
94  # =====
95  OUTPUT_DEFAULT_FORMAT=csv
96  OUTPUT_CSV_ENCODING=utf-8-sig
97  OUTPUT_EXCEL_SHEET_NAME=BaiduIndex
98  OUTPUT_FILE_NAME_TEMPLATE={task_type}_{timestamp}
99  OUTPUT_DIR=./output
100
101  # =====
102  # 阿里云OSS配置 (可选)
103  # =====
104  OSS_URL=
105  OSS_ENDPOINT=
106  OSS_ACCESS_KEY_ID=
107  OSS_ACCESS_KEY_SECRET=
108  OSS_BUCKET_NAME=
109  OSS_REGION=
```

!! 重要提示

MYSQL_PASSWORD: 替换为你在安装 MySQL 时设置的密码

三、安装后端依赖

确保在baidu-index-hunter-backend目录下。

代码块

```
1  # 创建 Python 虚拟环境 (推荐)
2  python3 -m venv baiduindexhunter
3
4  # 激活虚拟环境
5  # Windows (CMD)
6  baiduindexhunter\Scripts\activate
7
8  # Windows (PowerShell)
9  baiduindexhunter\Scripts\Activate.ps1
10
11 # macOS / Linux
12 source baiduindexhunter/bin/activate
13
14 # 安装依赖
15 pip install -r requirements.txt
16 # 如果安装速度慢,可以使用国内镜像
17 pip install -r requirements.txt -i https://pypi.tuna.tsinghua.edu.cn/simple
```

四、安装前端依赖

代码块

```
1  #进入前端目录
2  cd baidu-index-hunter-frontend
3  # 安装依赖
4  npm install
5  # 如果速度慢,可以使用国内镜像
6  npm config set registry https://registry.npmmirror.com
7  npm install
```

五、启动服务

1. 启动后端服务

代码块

```
1  # 打开一个命令行窗口, 进入后端目录
2  cd baidu-index-hunter-backend
3  # 激活虚拟环境
4  # Windows (CMD)
5  baiduindexhunter\Scripts\activate
6
7  # Windows (PowerShell)
8  baiduindexhunter\Scripts\Activate.ps1
9
10 # macOS / Linux
11 source baiduindexhunter/bin/activate
12
13 #启动后端
14 #python3 app.py
15 python app.py
```

看到以下输出表示启动成功:

代码块

```
1  > cd baidu-index-hunter-backend
2  > python3 app.py
3  2026-01-13 20:19:26.019 | INFO      | utils.logger:info:39 - 成功连接到MySQL和
  Redis
4  2026-01-13 20:19:26.697 | INFO      | logging:callHandlers:1706 - NumExpr
  defaulting to 8 threads.
```

```
5 2026-01-13 20:19:27.116 | INFO | utils.logger:info:39 - 已同步0个账号的使用量
数据
6 2026-01-13 20:19:27.168 | INFO | utils.logger:info:39 - 已更新Cookie列表, 共
104 个可用Cookie
7 2026-01-13 20:19:27.171 | INFO | utils.logger:info:39 - 请求频率限制器初始化:
间隔 0.1~0.3秒
8 2026-01-13 20:19:27.172 | INFO | utils.logger:info:39 - Cipher-Text.js加载成
功
9 2026-01-13 20:19:27.178 | INFO | utils.logger:info:39 - 百度指数爬虫初始化完成
10 2026-01-13 20:19:27.192 | INFO | utils.logger:info:39 - 已同步0个账号的使用量
数据
11 2026-01-13 20:19:27.235 | INFO | utils.logger:info:39 - 已更新Cookie列表, 共
104 个可用Cookie
12 2026-01-13 20:19:27.275 | INFO | utils.logger:info:39 - 默认配置初始化完成, 共
设置 22 项配置
13 2026-01-13 20:19:27.275 | INFO | utils.logger:info:39 - 开始同步区域数据到
Redis...
14 2026-01-13 20:19:27.286 | INFO | utils.logger:info:39 - 成功同步 34 条省份数
据到Redis
15 2026-01-13 20:19:27.298 | INFO | utils.logger:info:39 - 成功同步 369 条地级市
数据到Redis
16 2026-01-13 20:19:27.494 | INFO | utils.logger:info:39 - 成功同步 3376 条区域
关系数据到Redis
17 2026-01-13 20:19:27.503 | INFO | utils.logger:info:39 - 成功统计并同步 33 个
省份的城市数据到Redis
18 2026-01-13 20:19:27.613 | INFO | utils.logger:info:39 - 成功同步 104 个账号的
cookie数据到Redis, 其中 104 个可用
19 2026-01-13 20:19:27.653 | INFO | utils.logger:info:39 - 无需更新历史累计爬取数
据条数
20 2026-01-13 20:19:27.654 | INFO | utils.logger:info:39 - 没有待处理的任务
21 2026-01-13 20:19:27.656 | INFO | utils.logger:info:39 - 任务调度器启动成功, 工
作线程数: 16
22 2026-01-13 20:19:27.660 | INFO | utils.logger:info:39 - Task
cookie_status_check started: unknown
23 2026-01-13 20:19:27.661 | INFO | utils.logger:info:39 - 启动应用, 地址:
http://0.0.0.0:5001, 调试模式: 关闭
24 2026-01-13 20:19:27.661 | INFO | utils.logger:info:39 - 开始检查cookie状态...
25 2026-01-13 20:19:27.662 | INFO | utils.logger:info:39 - API文档地址:
http://0.0.0.0:5001/api/docs/
26 * Serving Flask app 'app'
27 * Debug mode: off
28 2026-01-13 20:19:27.669 | INFO | utils.logger:info:39 - 没有需要解封的cookie
29 2026-01-13 20:19:27.674 | INFO | utils.logger:info:39 - Cookie状态检查完成:
104/107 可用, 无需解封
30 2026-01-13 20:19:27.674 | INFO | utils.logger:info:39 - Task
cookie_status_check completed: unknown (0.012s)
```

```
31  2026-01-13 20:19:28.196 | INFO      | utils.logger:info:39 - 成功同步 104 个账号的
    cookie数据到Redis, 其中 104 个可用
32  2026-01-13 20:19:28.198 | INFO      | utils.logger:info:39 - 已重置Cookie缓存, 下
    次获取将重新加载
```

2. 启动前端服务

打开另一个命令行窗口:

代码块

```
1  #进入前端目录
2  cd baidu-index-hunter-frontend
3  #启动前端开发服务器
4  npm run dev
```

看到以下输出表示启动成功

代码块

```
1  > npm run dev
2
3  > baidu-index-hunter-frontend@0.0.0 dev
4  > vite
5
6
7  VITE v5.4.19  ready in 351 ms
8
9  ➔ Local:   http://localhost:5173/
10 ➔ Network: http://192.168.10.238:5173/
11 ➔ Network: http://198.18.0.1:5173/
12 ➔ press h + enter to show help
```

验证安装

1. 检查后端服务

在浏览器中访问:

- 后端首页: <http://localhost:5001/>
- API 文档: <http://localhost:5001/api/docs/>
- 健康检查: <http://localhost:5001/api/health>

如果能看到 JSON 响应,说明后端启动成功。

2. 检查前端服务

在浏览器中访问:

- 前端首页: <http://localhost:5173/>

如果能看到项目界面,说明前端启动成功。

3. 检查数据库连接

在前端界面中:

1. 点击左侧菜单的 "系统配置"
2. 如果能看到配置项,说明数据库连接成功

4. 添加 Cookie 并测试

1. 登录[百度指数](#)
2. 打开浏览器开发者工具 (F12)
3. 切换到 "Application" 或 "存储" 标签
4. 找到 Cookies,复制所有 Cookie
5. 在项目前端的 "Cookie 管理" 页面添加 Cookie
6. 在 "搜索指数采集" 页面测试一个简单关键词

!! 再次提醒: 本项目仅供学习交流使用,不可用于商业用途!