**项目简要**

目录

[1、 项目结构 1](#_Toc534378982)

[a. 工程图示例： 1](#_Toc534378983)

[b. 编译路径说明： 2](#_Toc534378984)

[c. 执行文件说明： 2](#_Toc534378985)

[d. 工作目录说明： 2](#_Toc534378986)

[2、 CollectorCore核心库 3](#_Toc534378987)

[3、 KeywordCollector关键字采集器 3](#_Toc534378988)

[a. 界面功能简介： 3](#_Toc534378989)

[b. 功能简述： 4](#_Toc534378990)

[4、 SnapshotCollector快照采集器 4](#_Toc534378991)

[a. 界面功能预览： 4](#_Toc534378992)

[b. 功能简述： 4](#_Toc534378993)

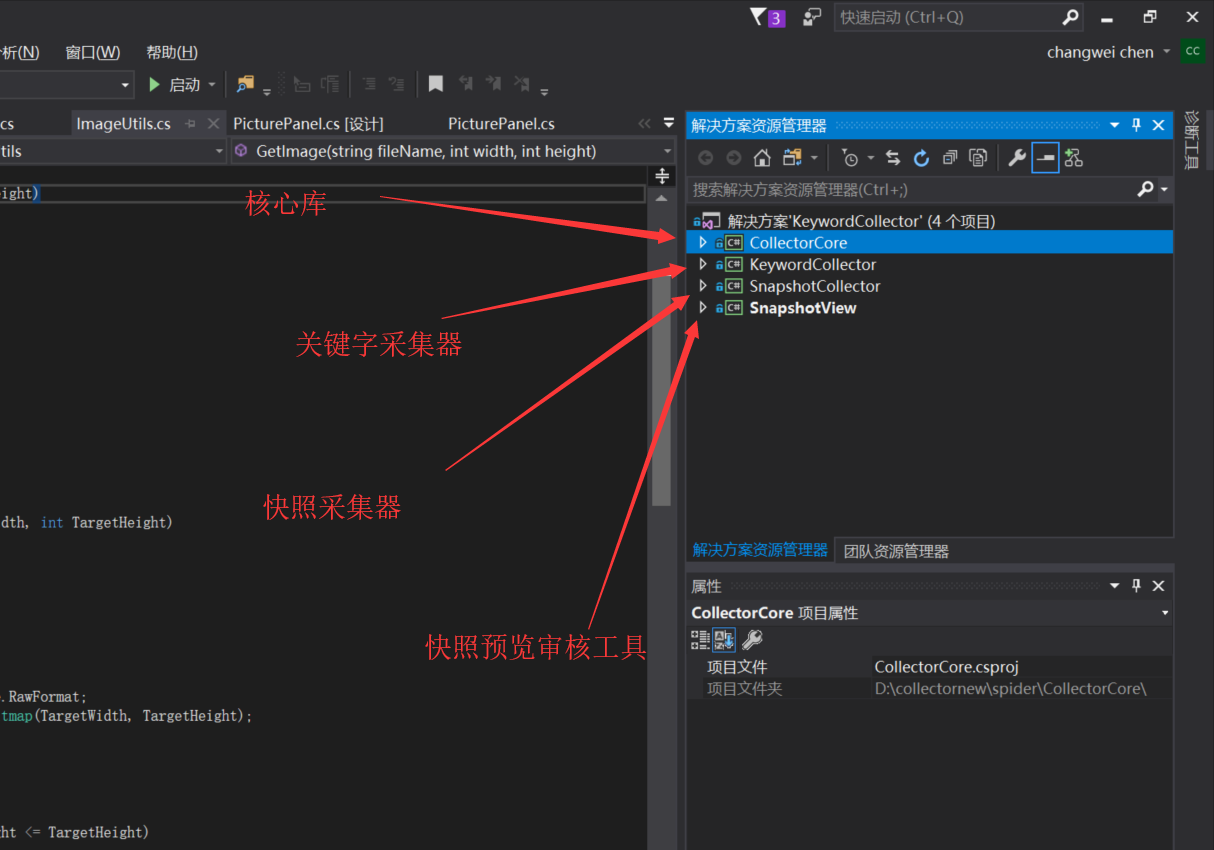
[5、 SnapshotView快照预览审核工具 5](#_Toc534378994)

[a. 界面功能预览： 5](#_Toc534378995)

[b. 功能简述： 5](#_Toc534378996)

## 项目结构

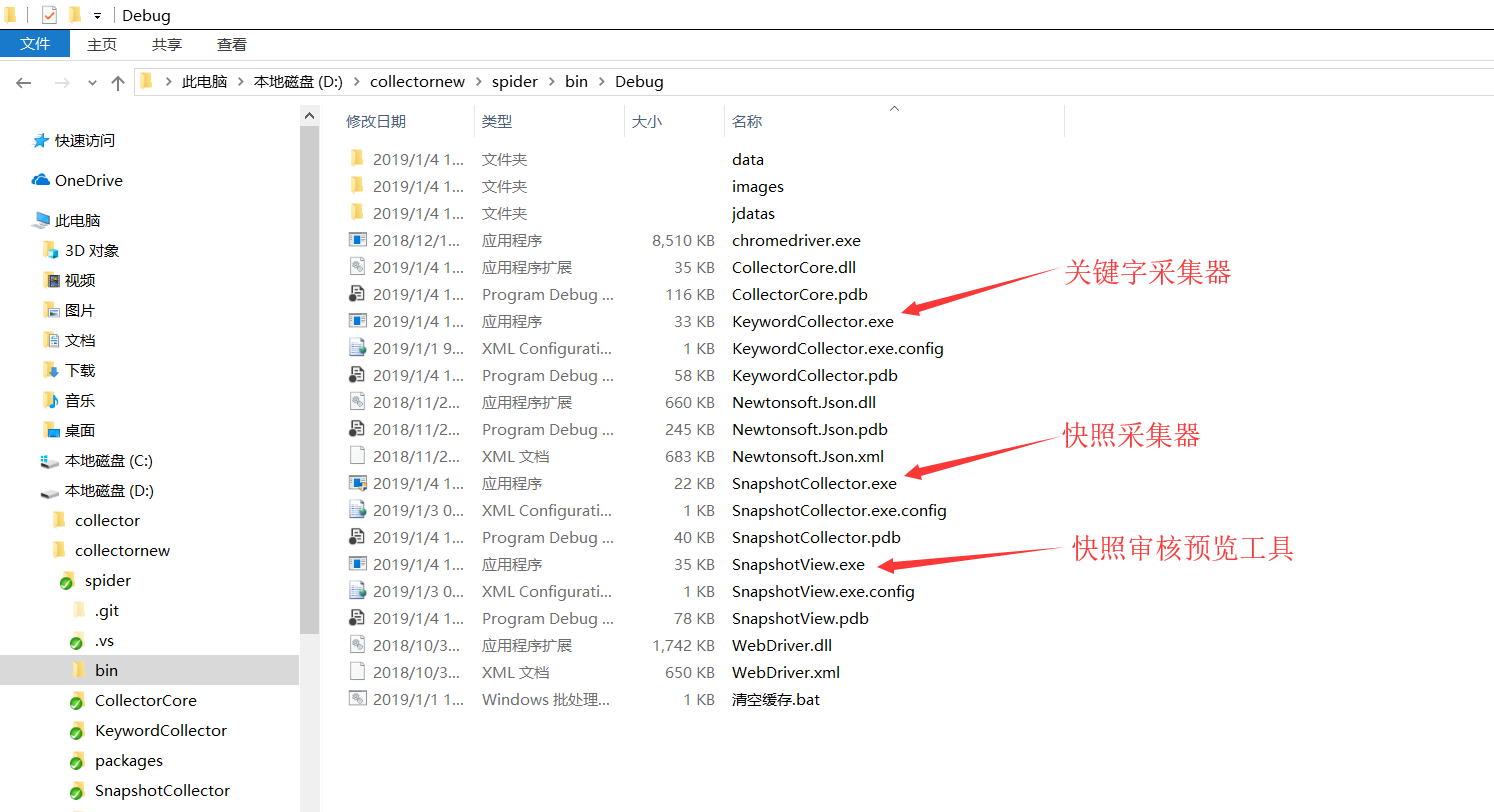
### 工程图示例：



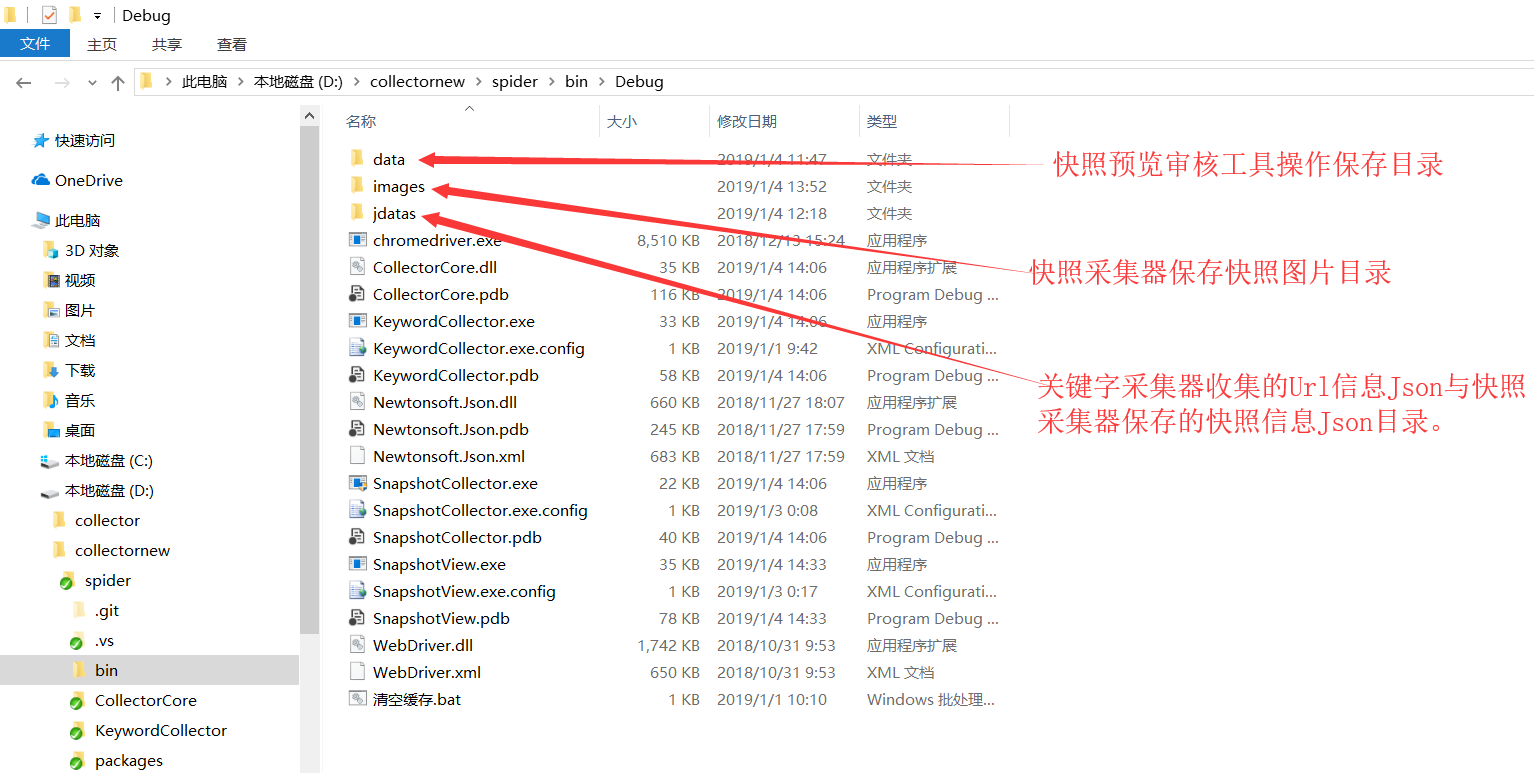
### 编译路径说明：

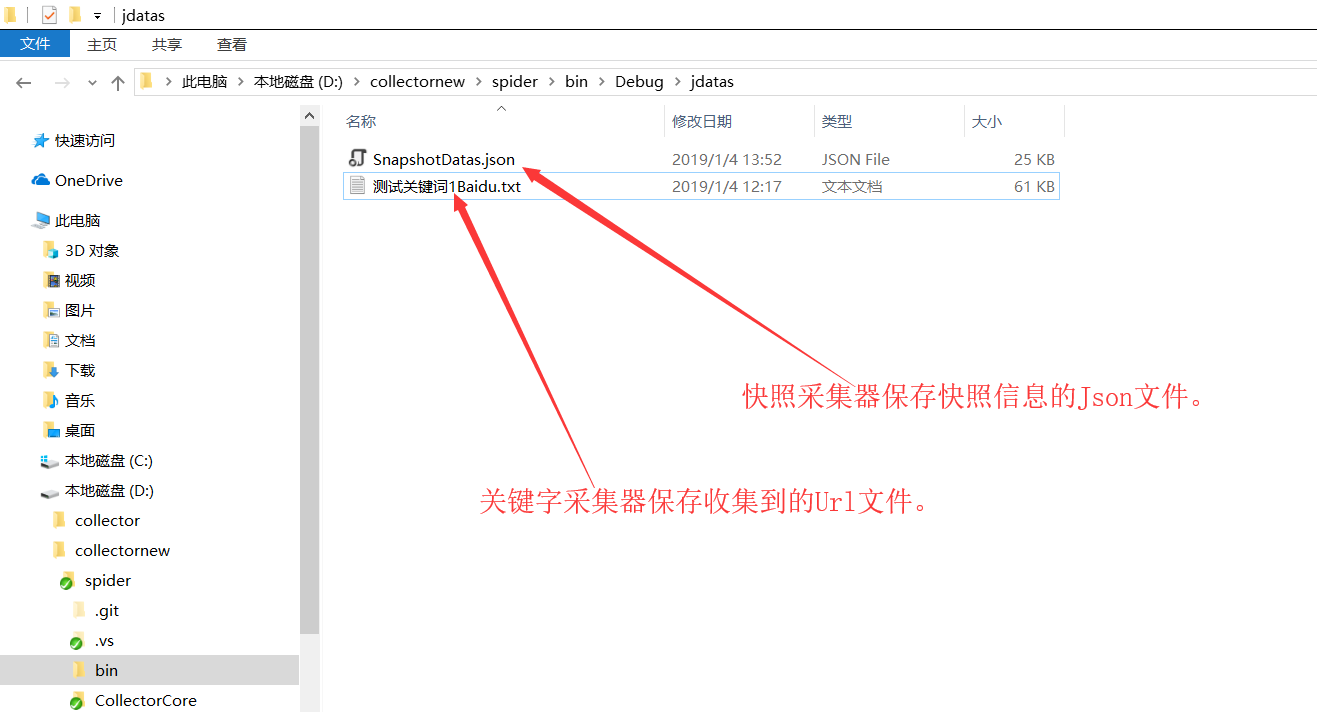
所有工程默认编译生成到..\bin\Debug\目录下，也就是解决方案根目录下。

### 执行文件说明：



### 工作目录说明：



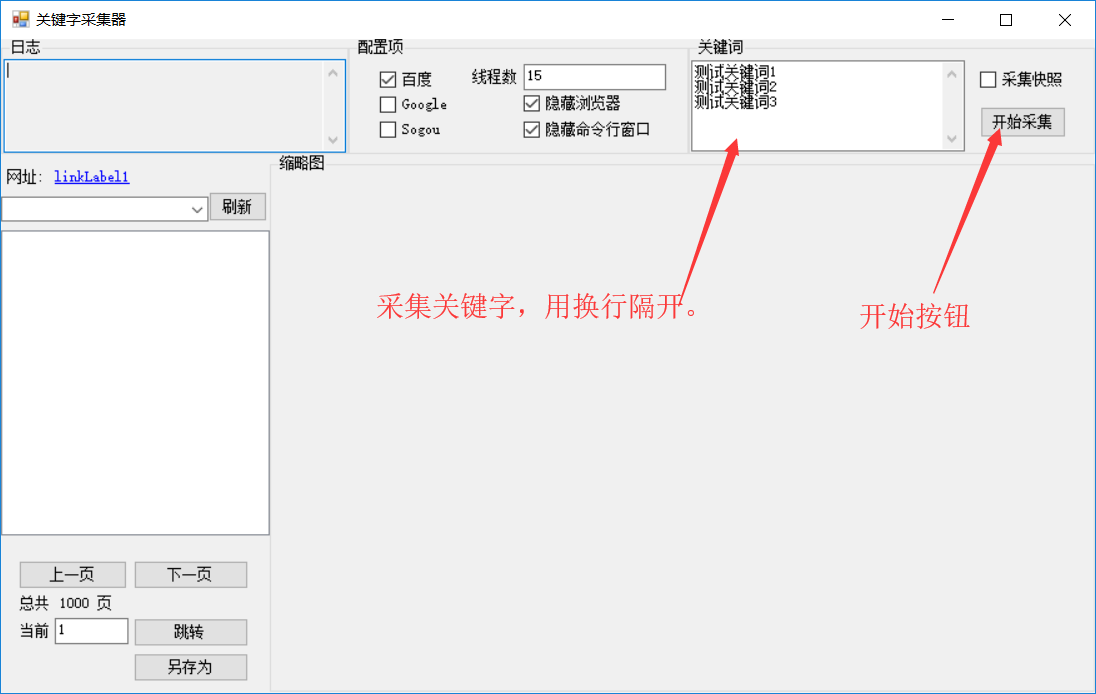


## CollectorCore核心库

项目核心代码库，包含搜索引擎页面爬虫、快照生成、文件数据库的核心块代码。

## KeywordCollector关键字采集器

### 界面功能简介：



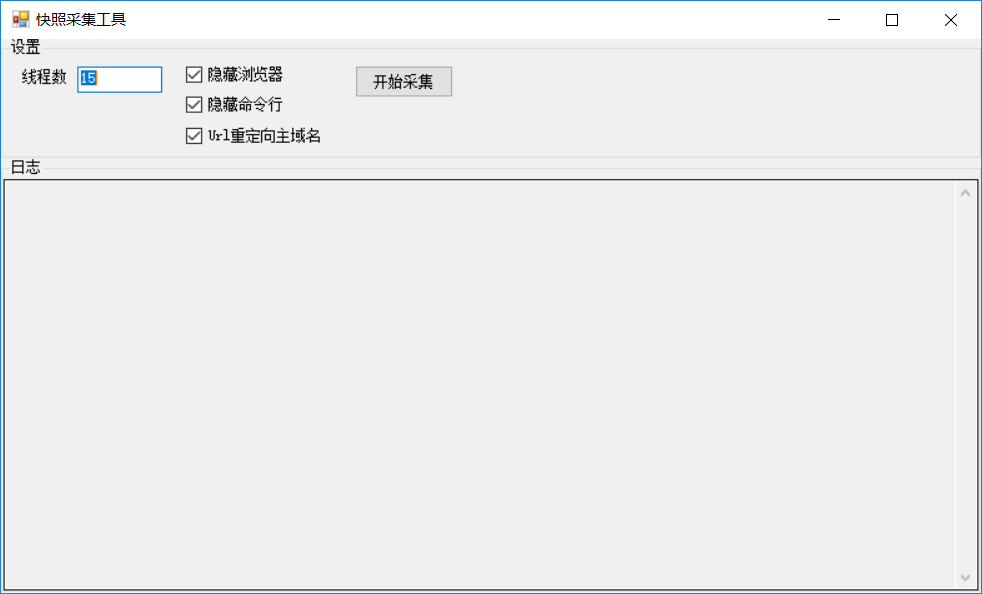
### 功能简述：

根据指定的一组关键字爬取百度、谷歌、搜狗搜索引擎的搜索结果URL，爬取的结果保存到\jdatas目录下，并以 {Keyword}{SearchEngineType}.txt 方式命名，比如：”足球网baidu.txt”。

保存的URL数据以换行符分割。

## SnapshotCollector快照采集器

### 界面功能预览：



### 功能简述：

1、点击开始采集按钮以后程序会搜索\jdatas目录所有的\*.txt文件，并把\*.txt文件所有的URL分割放到内存中，如果勾选Url重定向主域名程序会对所有的URL请求一次拿到重定向302的URL，再取出重定向302的URL。

2、生成的快照图片保存到\images目录下，另外没采集一个快照会把快照的信息写入到

\jdatas\ SnapshotDatas.json数据文件中。

## SnapshotView快照预览审核工具

### 界面功能预览：



### 功能简述：

程序启动以后会自动加载\jdatas\SnapshotDatas.json文件，该文件由快照采集器生成，保存的内容是URL与快照图片关联的信息。