**中国矿业大学计算机学院**

**2017级本科生实验报告**

课程名称 软件工程实践

实验题目 C#语言与集成开发环境

学生姓名 郭子杨

学 号 08173042

专业班级 计科2017-6班

任课教师 刘迎春

目录

[1.开发环境基本情况 3](#_Toc28356968)

[1.1.C#语言包版本： 3](#_Toc28356969)

[1.2.开发工具： 3](#_Toc28356970)

[2.C#介绍及语言编程规范 3](#_Toc28356971)

[6.1. 简介 3](#_Toc28356972)

[6.2. 命名规则 3](#_Toc28356973)

[6.3. 布局约定 3](#_Toc28356974)

[6.4. 参考文档 4](#_Toc28356975)

[3.开发环境操作介绍 4](#_Toc28356976)

[7.1. Vscode界面介绍 4](#_Toc28356977)

[7.2. 常用功能介绍： 5](#_Toc28356978)

[7.3. 文件常用操作： 7](#_Toc28356979)

[7.4. 编辑常用操作： 7](#_Toc28356980)

[7.5. 查看常用操作: 7](#_Toc28356981)

[7.6. 调试常用操作： 7](#_Toc28356982)

[7.7. 插件面板： 8](#_Toc28356983)

[7.8. 右键类或者方法： 9](#_Toc28356984)

[4.开源项目基本情况 10](#_Toc28356985)

[4.1.运行截图： 10](#_Toc28356986)

[4.2.项目说明 12](#_Toc28356987)

[4.3.功能架构 12](#_Toc28356988)

[4.4.第三方库 14](#_Toc28356989)

[5.开源项目需求分析 14](#_Toc28356990)

[6.开源项目软件结构 14](#_Toc28356991)

[6.1.软件类图： 14](#_Toc28356992)

[6.2.部分实例的时序图如下： 16](#_Toc28356993)

[6.3.项目内容图 17](#_Toc28356994)

[7.开源项目源码分析 18](#_Toc28356995)

[7.1.项目中各类的功能 18](#_Toc28356996)

[7.2.项目中的重要函数 18](#_Toc28356997)

[7.2.1．Void EMessageBoxType（） 19](#_Toc28356998)

[7.2.2．Void MakeIntent（Data，Action） 19](#_Toc28356999)

[7.2.3.void NewInstance() 19](#_Toc28357000)

[7.2.4.int FetchUrlConfig(url，config) 19](#_Toc28357001)

# 

# 

# 

# 

# 1.开发环境基本情况

1.1.C#语言包版本：1.21.9

1.2.开发工具：Visual Studio Code

Microsoft在2015年4月30日Build 开发者大会上正式宣布了 Visual Studio Code 项目：一个运行于 Mac OS X、Windows和 Linux 之上的，针对于编写现代 Web 和云应用的跨平台源代码编辑器。

VS vode特点：

* 开源，免费；
* 自定义配置
* 集成git
* 智能提示强大
* 支持各种文件格式（html/jade/css/less/sass/xml）
* 调试功能强大
* 各种方便的快捷键
* 强大的插件扩展

# 2.C#介绍及语言编程规范

## 简介

C#是微软推出的一种基于.NET框架的、面向对象的高级编程语言。C#以.NET框架类库作为基础，拥有类似Visual Basic的快速开发能力。C#由安德斯·海尔斯伯格主持开发，微软在2000年发布了这种语言，希望借助这种语言来取代Java。C#已经成为ECMA国际和国际标准组织的标准规范。

## 命名规则

在不包括 using 指令的短示例中，使用命名空间限定。 如果你知道命名空间默认导入项目中，则不必完全限定来自该命名空间的名称。 如果对于单行来说过长，则可以在点 (.) 后中断限定名称。

你不必更改通过使用 Visual Studio 设计器工具创建的对象的名称以使它们适合其他准则。

## 布局约定

使用默认的代码编辑器设置（智能缩进、4 字符缩进、制表符保存为空格）。

每行只写一条语句。

每行只写一个声明。

如果连续行未自动缩进，请将它们缩进一个制表符位（四个空格）。

在方法定义与属性定义之间添加至少一个空白行。

注释约定：

将注释放在单独的行上，而非代码行的末尾。

以大写字母开始注释文本。

以句点结束注释文本。

在注释分隔符 (//) 与注释文本之间插入一个空格。

## 参考文档

微软C#指南

<https://docs.microsoft.com/zh-cn/dotnet/csharp/programming-guide/inside-a-program/coding-conventions>

# 3.开发环境操作介绍

## Vscode界面介绍

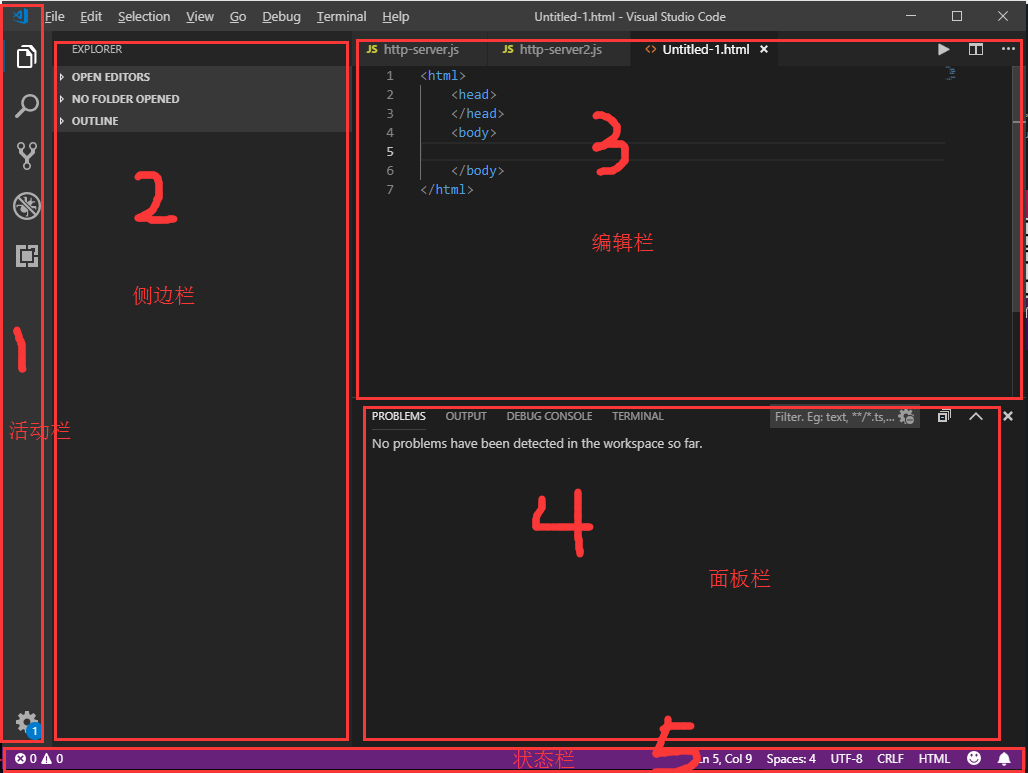


图3-1 Vscode界面

①：活动栏从上到下依次为，打开侧边栏，搜索，使用git，debug，使用插件

②：侧边栏，新建项目文件和文件夹

③：编辑栏，编写代码的区域

④：面板栏，从左到右依次为，问题，输出，调试栏，终端（terminal），最重要的是terminal，用来输入相关命令



图3-2面板栏上部

⑤：状态栏，点击该区域可以调出面板栏

⑥：需要注意的为下图红框所示，分别表示鼠标光标所在位置和tab缩进字符，这里为缩进4个字符



图3-3状态栏

## 常用功能介绍：



图3-4 文件选项 图3-5 编辑选项



图3-6 查看选项 图3-7 调试选项



图3-8终端选项

## 文件常用操作：

新建project，导入已存在的资源project，从版本控制库导入工程，新建Module，导入已存在的资源Module

打开文件：打开本地的文件或工程

打开最近文件：打开最近已导入过的工程

保存文件：将文件保存至本地

将文件夹添加到工作区：将文件加入工作区的资源管理器

## 编辑常用操作：

查找：查找指定内容

替换：替换制定内容

撤销：撤销操作

恢复：恢复操作

## 查看常用操作:

输出：输出并查看

资源管理器：查看资源管理器

源代码管理：打开源代码管理器

调试控制台：打开调试控制台，可以逐步调试

调试：调试文件

## 调试常用操作：

启动调试：开启调试，并切换调试界面

打开配置：打开调试配置

添加配置：增加调试配置

新建断点：建立断点

切换断点：切换至某个调试断点

## 插件面板：

插件管理和下载：

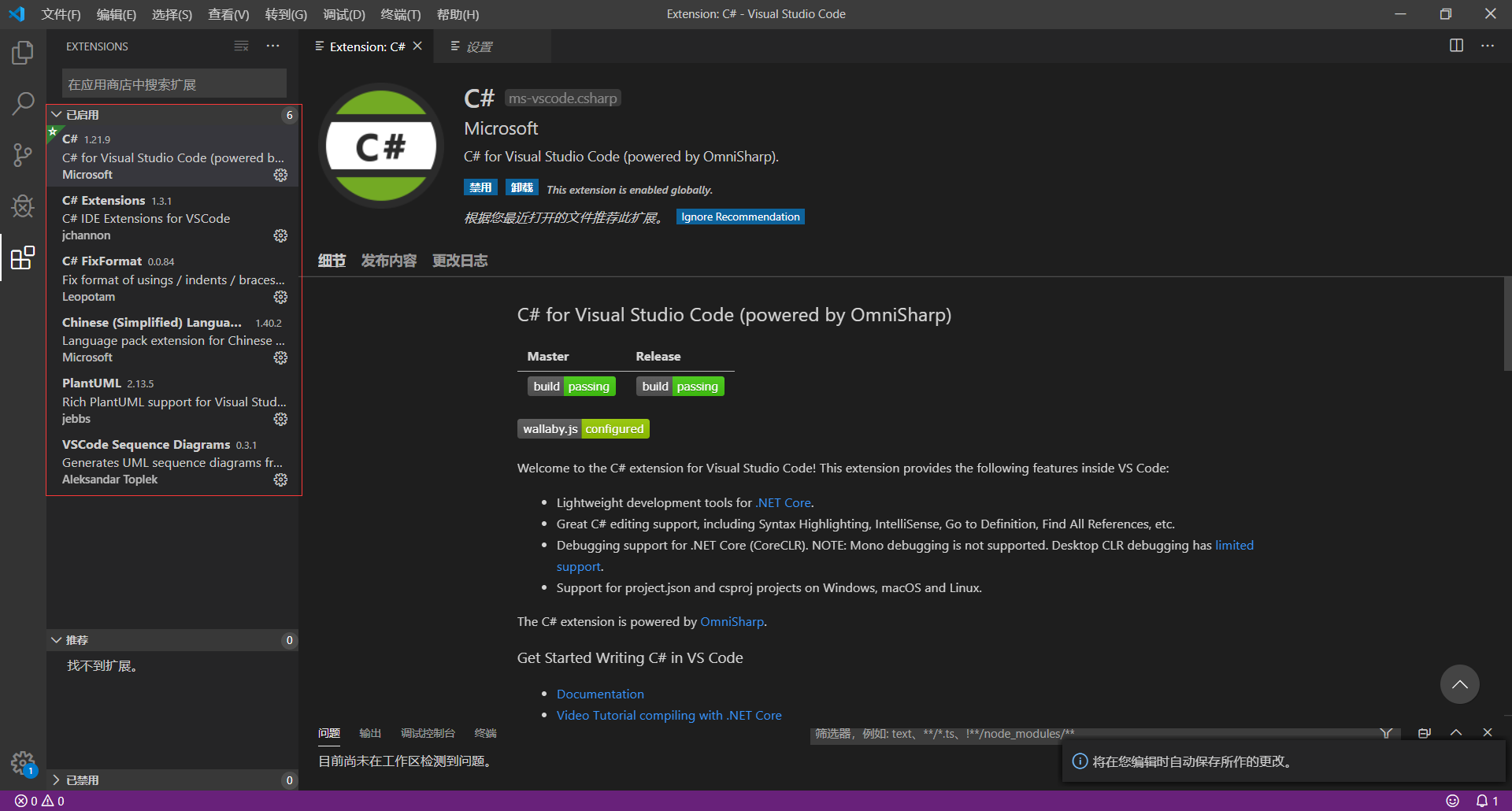


图3-9插件下载

1. C#语言包：进行C#编程必需的语言包
2. C#扩展工具：提供C#扩展集成开发环境工具
3. C#格式化工具：C#自动排版的工具
4. 中文语言补丁：是Vscode显示中文
5. PlantUML：用于类图绘制
6. Sequence Diagrams：用于时序图的绘制

## 右键类或者方法：



图 3-10生成时序图



图3-11生成类图

在代码中点击右键，可以生成时序图。在资源管理器中点击右键，可以生成类图。

# 4.开源项目基本情况

开源项目： C#爬虫项目（[CSharpCrawler](https://github.com/zhaotianff/CSharpCrawler/tree/master/CSharpCrawler)）

来自于：<https://github.com/zhaotianff/CSharpCrawler>

4.1.运行截图：

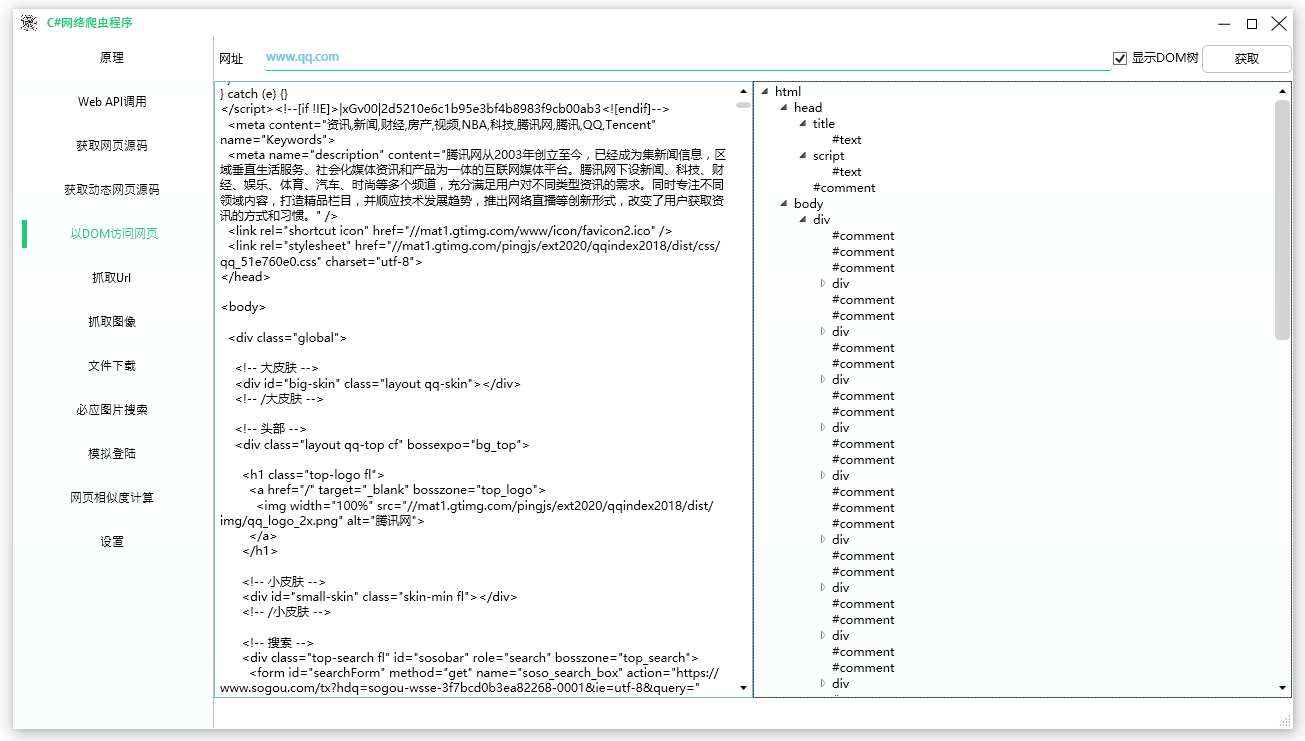


图4-1以DOM访问网页

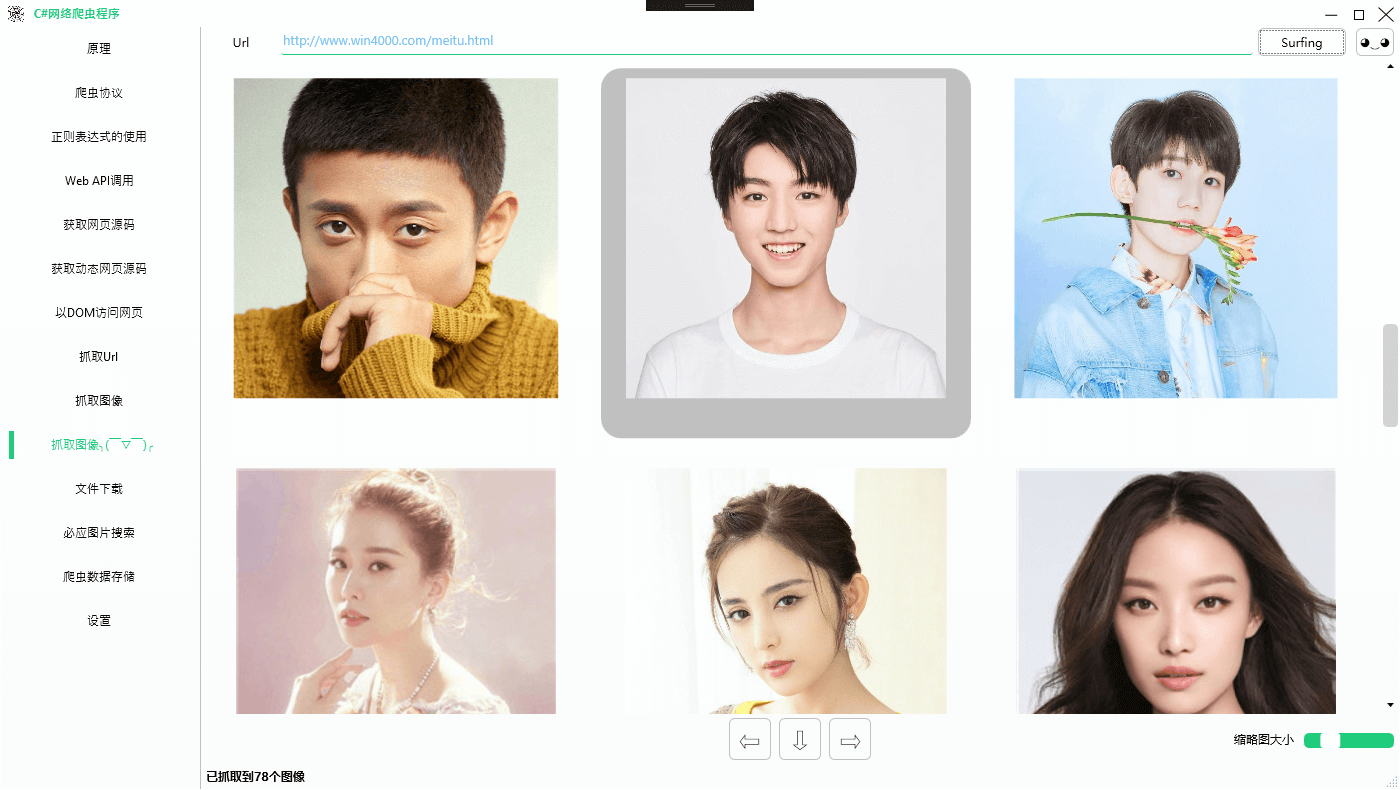


图4-2抓取图像

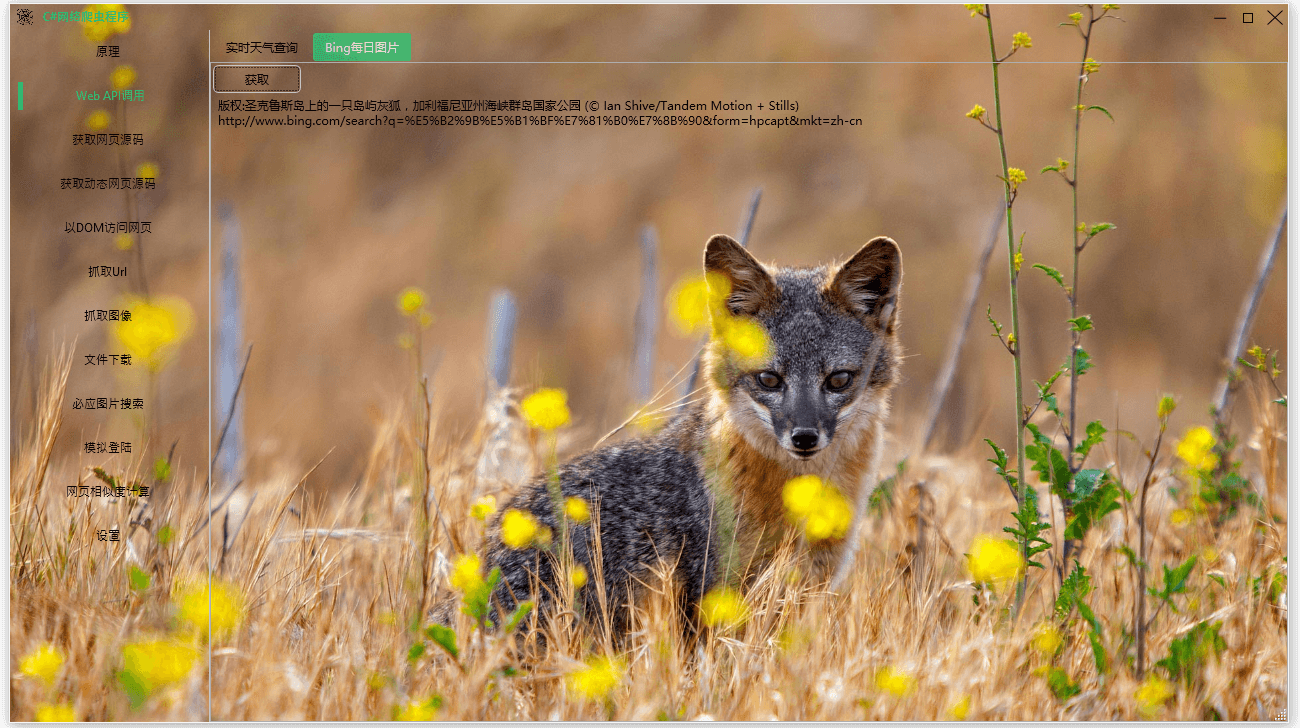


图4-3Web API调用

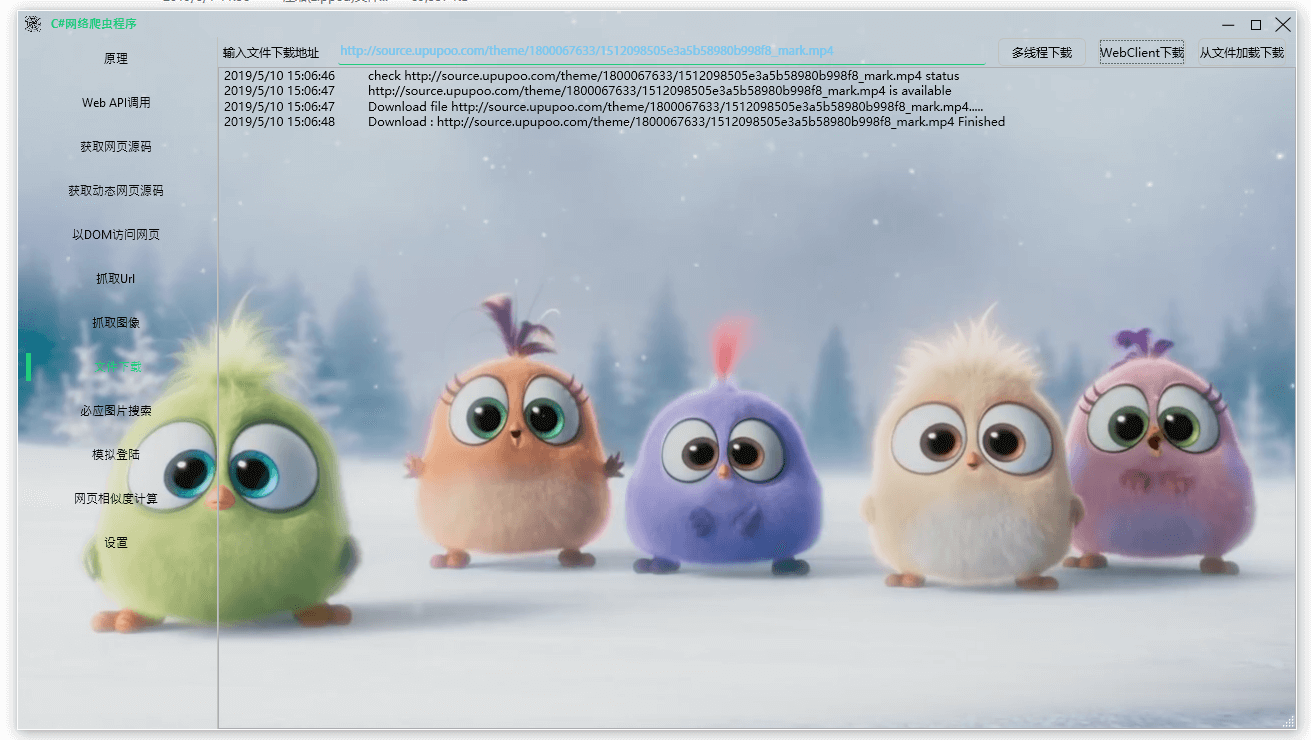


图4-4文件下载

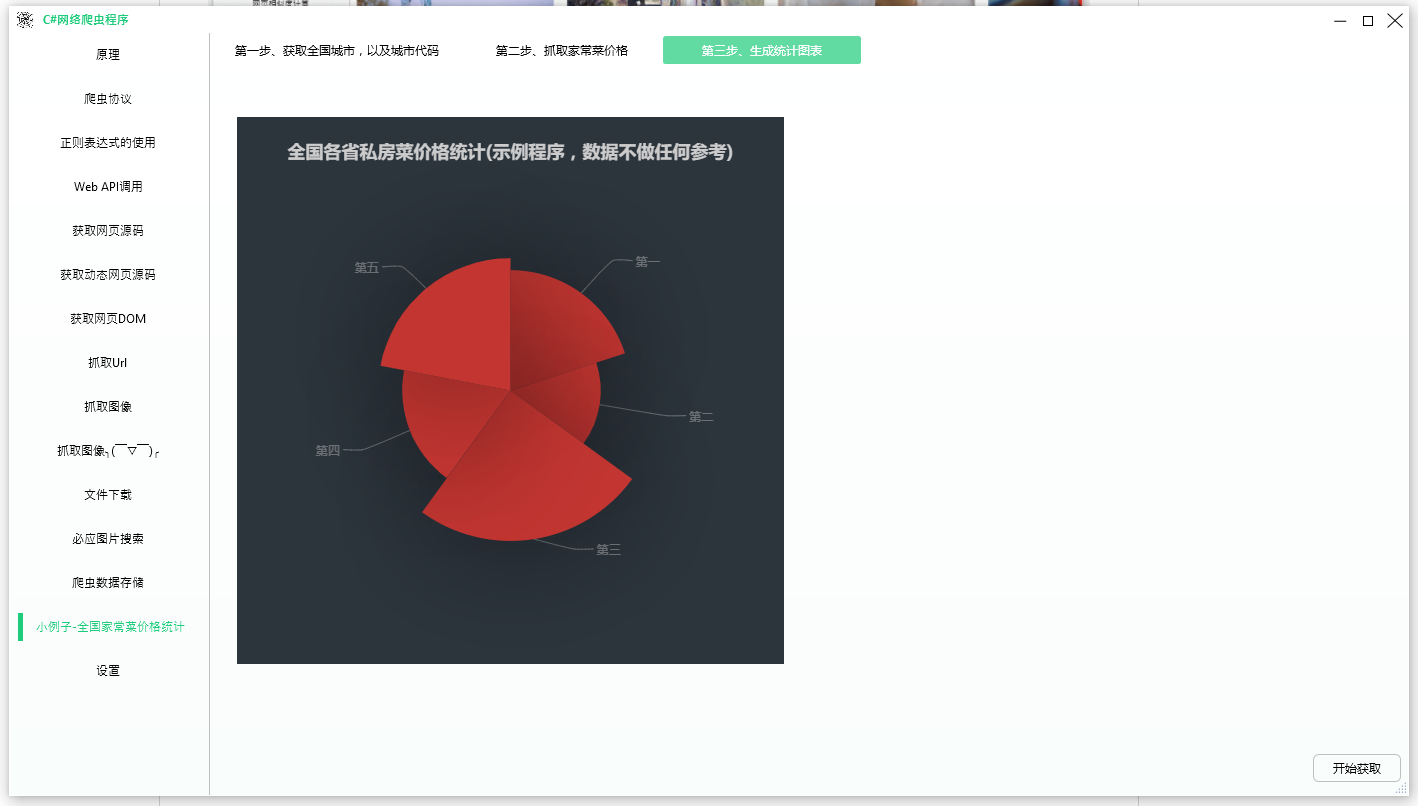


图4-5全国家常菜价统计

## 4.2.项目说明

网络爬虫（又称为网页[蜘蛛](https://baike.baidu.com/item/%E8%9C%98%E8%9B%9B/8135707)，网络机器人，在[FOAF](https://baike.baidu.com/item/FOAF)社区中间，更经常的称为网页追逐者），是一种按照一定的规则，自动地抓取[万维网](https://baike.baidu.com/item/%E4%B8%87%E7%BB%B4%E7%BD%91/215515)信息的程序或者脚本。另外一些不常使用的名字还有[蚂蚁](https://baike.baidu.com/item/%E8%9A%82%E8%9A%81/9770178)、自动索引、模拟程序或者[蠕虫](https://baike.baidu.com/item/%E8%A0%95%E8%99%AB/4454380)。

本项目是一个基于C#语言的爬虫，具有轻量化的特点。

## 4.3.功能架构

* 基础知识
  + [爬虫基础知识](https://github.com/zhaotianff/CSharpCrawler/blob/master/CSharpCrawler/PrerequisiteKnowledge.md)
* 网页抓取原理
  + 使用套接字来获取网页源码
* 爬虫协议
  + [爬虫协议介绍，以及它的语法规则;](https://github.com/zhaotianff/CSharpCrawler/blob/master/CSharpCrawler/RobotsExclusionProtocol.md)
  + C#如何获取网站的爬虫协议;
  + C#中如何解析爬虫协议;
* [正则表达式的使用](https://github.com/zhaotianff/CSharpCrawler/blob/master/CSharpCrawler/%E6%AD%A3%E5%88%99%E8%A1%A8%E8%BE%BE%E5%BC%8F.md)
  + 正则表达式的基础知识和基本使用;
  + 正则表达式中的分组构造;
  + 常用匹配模式;
* WebAPI调用
  + 获取实时天气
    - 调用中国天气网公开API接口来获取天气
  + 获取Bing每日图片
    - 调用cn bing API接口来获取Bing每日图片
* 抓取网页
  + HttpWebRequest类的使用
  + 获取指定url的IP地址
  + 获取指定url的网页头信息
  + 如何从网页源码中提取页面的编码
* 抓取动态网页
  + 使用[CEFSharp](https://github.com/cefsharp/CefSharp)来抓取动态网页
  + 使用WebBrowser(IE)来抓取动态网页
  + 使用[Selenium](https://github.com/SeleniumHQ/selenium)来抓取动态网页
* 获取网页DOM
  + 使用[HtmlAgilityPack](https://github.com/zzzprojects/html-agility-pack)来获取网页的DOM结构
* Url抓取(当Url太多时，UI会卡)
  + 抓取指定网址的全部链接
  + 通过指定深度，抓取子网页的全部链接
  + 动态网页链接抓取
  + 限定抓取当前页面的子链接
* 图片抓取(更新中)
  + 抓取指定url页面中的图片，通过配置url的页码规则，来进行翻页。
  + 自动获取下一页
* 文件下载
  + 使用WebClient类下载文件
  + 多线程下载文件
  + 从文件加载批量下载
* 必应图片搜索(仅供学习，请勿用途其它用途)
  + 实现必应图片搜索的功能
  + 翻页及优化(待更新)
* 爬虫数据存储
  + Berkeley DB
    - 介绍BerkeleyDB以及使用方式
  + SQLite
    - 介绍SQLite以及使用方式
* 小例子-全国家常菜价格统计
  + 获取全国城市，以及城市代码
  + 抓取家常菜价格
  + 生成统计图表
* 模拟登录并获取登录后的内容
  + 使用Cookie(实现中)
  + [使用Selenium](https://github.com/zhaotianff/CSharpCrawler/blob/master/CSharpCrawler/Selenium.md)(实现中)
    - 说明：示例程序使用的是EdgeDriver，所以需要Windows10系统，如果需要其它浏览器Driver,可自行修改。
    - 测试系统：Windows 10 1703 Edge 15.15063.0，如果Edge驱动版本不一致，需要手动更新至对应的版本。
* 抓包工具使用
  + [Fiddler](https://github.com/zhaotianff/CSharpCrawler/blob/master/CSharpCrawler/AnalysisPacket_Fiddler.md)

## 4.4.第三方库

* [CefSharp](https://github.com/cefsharp/CefSharp)
* [HtmlAgilityPack](https://github.com/zzzprojects/html-agility-pack)
* [Oracle Berkeley DB](https://www.oracle.com/database/technologies/related/berkeleydb.html)
* [SQLite](https://www.sqlite.org/index.html)
* [Json.NET](https://github.com/JamesNK/Newtonsoft.Json)
* [Selenium](https://github.com/SeleniumHQ/selenium)

# 5.开源项目需求分析

* 使用CSS选择器和XPath选取元素
* 抓包工具Charles的使用
* 使用抓包工具分析网站接口
* 使用抓包工具分析APP接口
* 将网页保存为图片/PDF
* 验证码识别(字符验证码，滑块验证码)
* BloomFilter算法
* NLP基础
* 中文分词
* Lucene.net使用
* 优先级队列实现
* 基本爬虫架构
* 分布式爬虫架构
* 抓取豆瓣书评
* 当抓取的数据到非常大的的数量级时该怎么处理
* 使用代理

# 6.开源项目软件结构

## 6.1.软件类图：

WindowCommand：



图6-1 WindowCommand的类图

BingImageInfo：

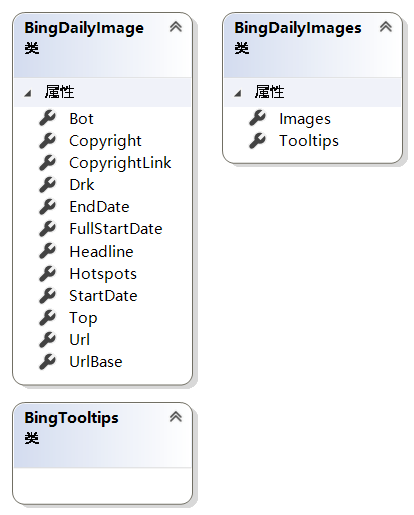


图6-2 BingImageInfo的类图

ConfigStruct：

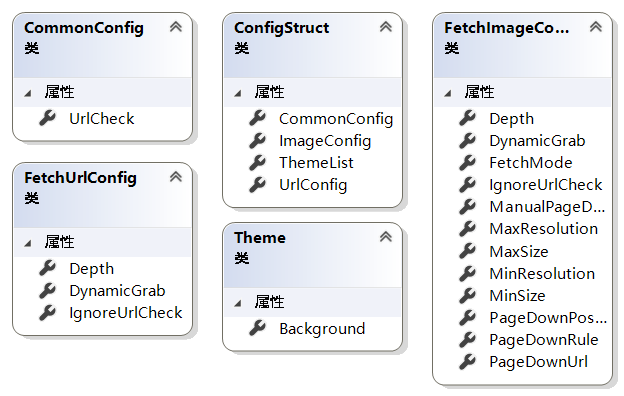


图6-3 ConfigStruct的类图

## 部分实例的时序图如下：

WebUtil\_GetHeader：

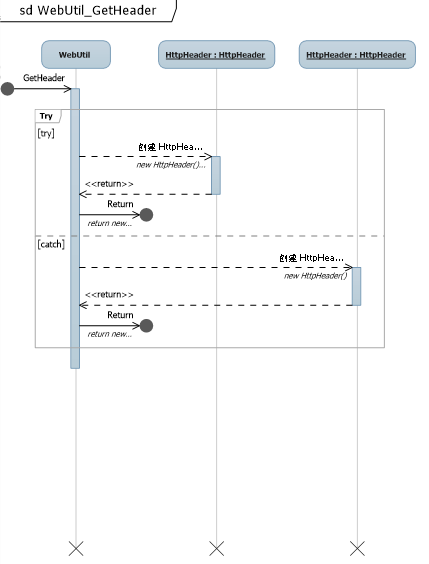


图6-4WebUtil\_GetHeader的时序图

HtmlAgilityPackUtil：

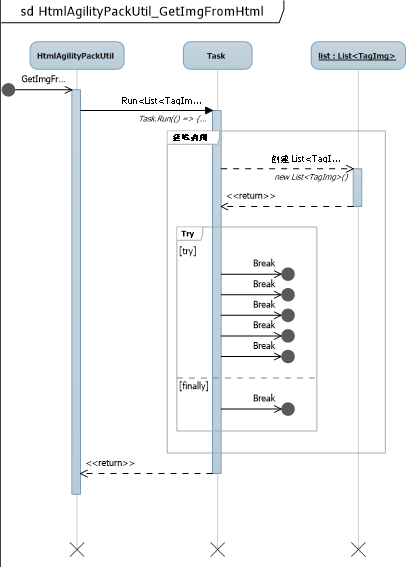


图6-5HtmlAgilityPackUtil的时序图

## 项目内容图

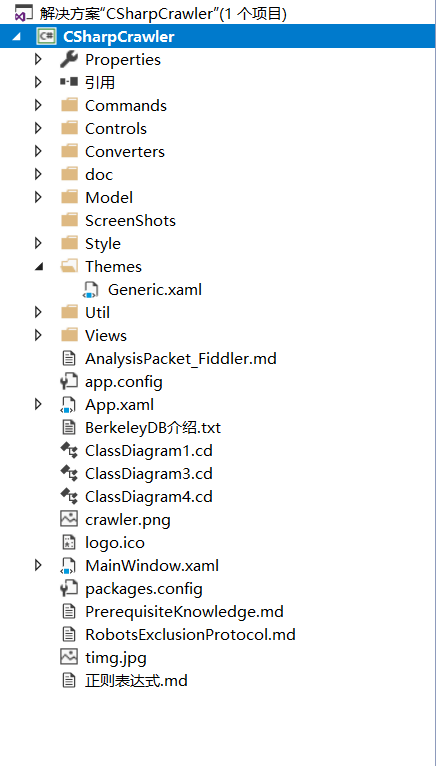


图6-6项目内容图

# 7.开源项目源码分析

## 7.1.项目中各类的功能

类ListImage负责列出网页抓取的图片

类CityWeatherCode负责列出城市的天气

类BingImageInfo负责列出必应的图片

类EMessageBox负显示抓取信息

类HtmlStruct负责显示HTML网页结构

类NodeStruct负责显示借点结构

类UrlStruct负责建立Url链接

类TreeNode负责建立节点树

## 7.2项目中的重要函数

### 7.2.1．Void EMessageBoxType（）

开发过程中难免要使用到消息框，然而系统提供的MessageBox却难以满足许多需求。一、MessageBox的背景颜色无法更改，这就无法满足需求要求的消息框颜色。二、MessageBox的提示形式过于单一，难以满足包含ListPicker、CheckBox等方式的弹出框，而CustomMessageBox的模版中包含了ContentPresenter，可以承载各种各样的内容控件。在多方面考虑之后，于是决定选择Toolkit的CustomMessageBox作为应用的弹出提示框。

### 7.2.2．Void MakeIntent（Data，Action）

这个命令可以在终端构造特定格式的Intent并且发送这个广播，这个命令的参数很丰富，可以满足大部分Intent构造需求。Intent是一个任务或功能的抽象，它会被APP在未来某个时刻执行。它告诉APP须要做的事情。Actions:也就是Intent须要做的事情。在软件中用于打开一个URL，编辑某些数据。一个Action是一个简单字符串常量描写叙述应该完毕的事情。Data:即Intent需用使用的资源。在Android中它会用一个URI(唯一资源标识符)或者Uri对象来表示。数据类型须要依据Action而变。将 Action和Data组合在一起。Android 就能够知道这个Intent是干什么的以及用什么来干的问题。

### 7.2.3.void NewInstance()

在开发中，我们经常可以看到一些接口的设计，实现接口的类加载一般都会采用newInstance。先调用Class加载方法加载某个类，然后实例化。可以在调用class的静态加载方法forName时获得更好的灵活性，提供给了一种降耦的手段。

### 7.2.4.int FetchUrlConfig(url，config)

fetch常用的接口为fetch(url, config)， 其中url为请求的地址，config为请求的配置对象包括配置请求方法，请求头等或者fetch(request)，其中 request为Request对象，可以通过new方法构造。