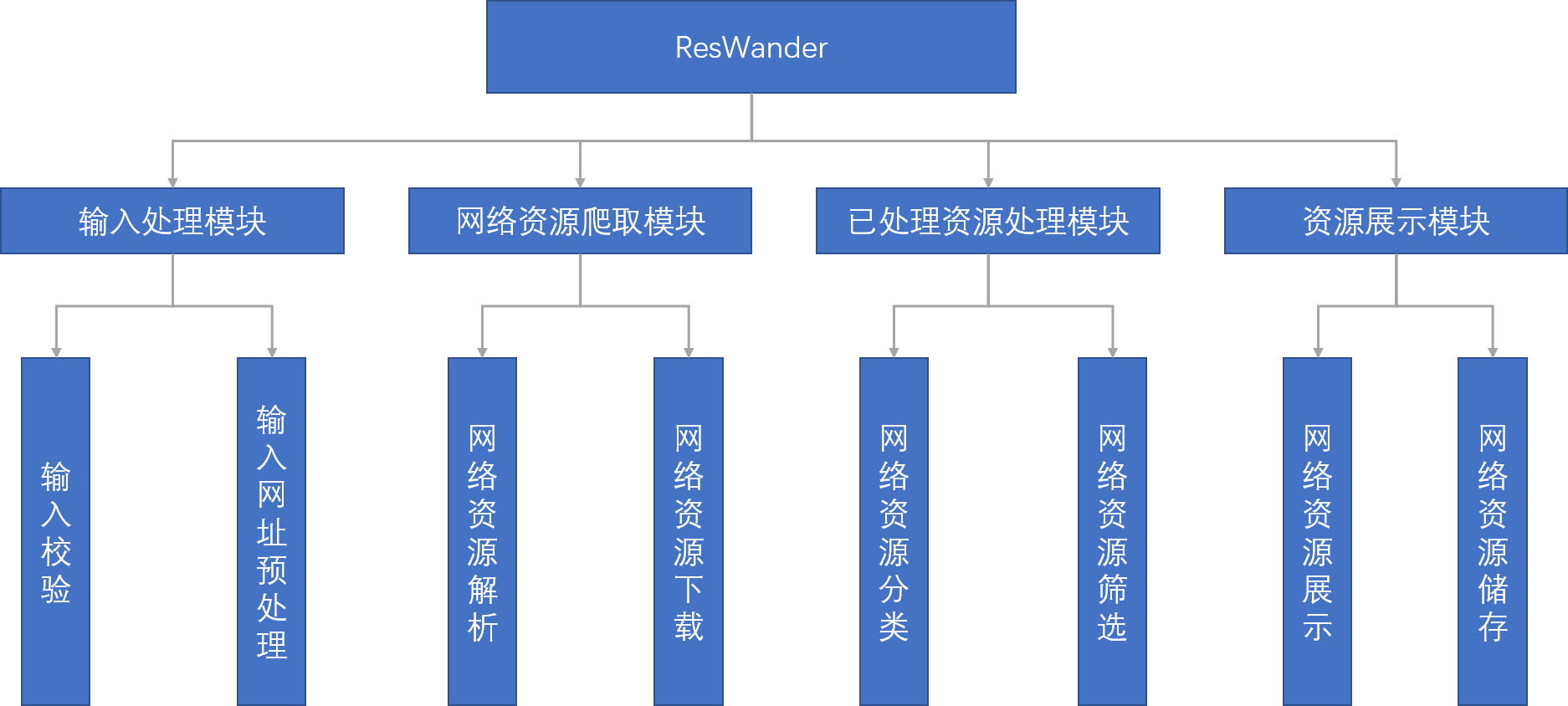
# 3.概要设计

## 功能模块设计



### 1.输入处理模块

输入：用户将爬取的网址、用户对爬取资源的要求(大小、类型)

输出：合法的规范化的网址，用户对爬取资源的要求

功能概述：提示用户输入待爬取的网页，以及对于爬取资源的要求，检验用户输入网址的合法性并作一定的加工处理，输入合法且规范化的网址，以及用户对爬取资源的要求。

#### 输入校验

输入：网站网址字符串

输出：合法的网址字符串

功能概述：根据用户输入，判断网址的字符串的合法性，若网址字符串合法，则输出网址输出，否则提示用户再次输入网址，直到输入合法。

#### 网址预处理

输入：合法的网址字符串

输出：规范化的网址字符串

功能概述：通过合法的网址字符串进行一定的处理，例如：加上：“http：//”等等，输出一个标准化、规范化的网址字符串，方便爬取函数的使用。

### 网络资源爬取模块：

输入：网址，爬取的资源类型

输出：当前网站上指定类型的资源

功能概述：利用规范化的网址在特定网站上进行资源爬取，并所有指定类型的资源下载下来。

#### 网络资源解析

输入：网址、爬取的资源类型

输入：目标资源的URL

功能概述：利用规范化的网址输入，对网页的html代码进行解析，找到目标资源的url，并将其输出。

#### 网络资源下载

输入：目标资源url

输出：网络资源实体

功能概述：利用每个资源对应的url，将目标资源实体下载到本地。

### 已爬取资源处理模块

输入：下载至本地的网络资源实体，用户对资源的要求

输出：分类清晰、满足用户需求的网络资源实体

功能概述：对原始的，刚下载完毕的资源进行分类整理，并根据根据用户的要求，对资源进行一定程度的筛选，最后将分类、加工完成的资源展示给用户

#### 网络资源分类

输入：网络资源实体

输出：不同类别的网络资源实体

功能概述：对输入的网络资源实体进加工，根据网络资源的属性(文件类型、文件大小等)对网络资源分类，并输入分类好的网络资源

#### 网络资源筛选

输入：网络资源实体，用户对资源的要求

输出：进行过筛选的网络资源实体

功能概述：根据输入的用户要求，对资源进行筛选，过滤掉低质资源（例如：文件大小过小）以及用户不需要的资源，将精简过的资源输出。

### 资源展示模块

#### 网络资源展示

输入：网络资源

输出：图形化界面

功能概述：生成图形化界面，并将输入的网络资源展示在图形化界面上，与用户进行交互，供用户查看，并下载图片。

#### 网络资源储存

输入：待存储的网络资源实体，存储路径

输出：存储至本地的网络资源实体

功能概述：用户可以通过图形化界面选择需要进行存储的资源以及资源保存位置，系统可以将指定网络资源存放到操作系统的指定位置。

## 核心类图

