deva

发布 *0.92*

2023 年 09 月 17 日

Contents

1 Indices and tables **1**

2 deva 解决的问题 **3**

2.1 案例 1: 数据输入爬虫解决的痛点 3

3 解决 py 入门程序员编程的基础问题 **5**

4 数据和空间问题：store **7**

4.1 解决结构化数据存储难题 7

4.2 解决中文数据对存储搜索功能 7

5 数据结构管理问题 NS NB NT **9**

6 时间时机问题：when **11**

7 多任务调度问题 catch，filter： **13**

8 多个 py 程序协作问题 NT，bus **15**

9 开发速度问题：Pipe **17**

10 后台程序执行过程的可视化 webview **19**

11 jupyter 中调试编写难题 **21**

12 和操作系统交互等一些难点 **23**

13 流程设计和编排和计算难题 **25**

14 实时持续产生数据的计算 **27**

CHAPTER 1

Indices and tables

- genindex
- modindex
- search

CHAPTER 2

deva 解决的问题

2.1 案例 1: 数据输入爬虫解决的痛点

1. 框架难，很多框架都有自己目录结构，代码逻辑分布在不同文件内，浏览修改以及记忆框架结构都比较难。deva 是库，在一个文件页面内展示完所有逻辑，阅读修改运行以及和其他代码集成都方便。
2. 分析页面难，很多教程用 bs，用 re。deva 除了有 xpath cssselector 等传统提取方法外，还有 find 和 search 方法，理解起来比 xpath 之类简单直观，还有 absolutelink 的提取，符合使用习惯
3. 整站抓取难。数据和导航链接的区分，循环 stream，去重复，很多实用功能
4. 并发加速难，普通教程默认都是 requests，堵塞的速度慢，复杂教程则引入 async 概念，把事情搞复杂。deva 只需要提交 request 参数，默认使用异步 http 库，用户对异步使用无感知。
5. 限速爬取难，有些网站做了反爬，需要限速，普通爬虫用 sleep，很容易让整个程序卡住，停爬的时候分析提取也停了，deva 自带 ratelimit，限速不要太方便

CHAPTER 3

解决 py 入门程序员编程的基础问题

CHAPTER 4

数据和空间问题：store

4.1 解决结构化数据存储难题

1. 读写文件保存数据，使用时找到数据很不方便
2. 使用 pickle 存储对象，何时保存何时加载控制难度大
3. 使用 mysql 需要懂 sql，建立连接执行 sql 也复杂
4. 使用 mongo 等 nosql 数据库，需要安装部署数据库程序，安装和数据备份运费运维复杂度升高
5. 实用内置 sqlite，不支持多线程

6. 每次写入和读取繁琐 deva 使用 sqlite 做底层，封装多线程安全，不需要考虑文件名，密码等，只需要记得表名即可。可直接存储大部分对象，generator 不可以，无需显式读写，使用和使用字典一样，批量存储一些数据比字典还方便

4.2 解决中文数据对存储搜索功能

CHAPTER 5

数据结构管理问题 NS NB NT

全局命名对象，防止重复对象，多模块中使用相同对象

CHAPTER 6

时间时机问题：when

解决和时间时机有关任务执行的痛点 没有内置的合适的间隔执行没有定时定点执行函数的功能程序退出时需要关闭或者保存到内容

CHAPTER 7

多任务调度问题 catch，filter：

同时处理多个事情的程序编写难题 异步任务的执行事件驱动信号驱动编程

CHAPTER 8

多个 py 程序协作问题 NT，bus

跨进程的数据或对象传递多个进程相互信号调度的问题

CHAPTER 9

开发速度问题：Pipe

解决自己写函数的问题 多个括号嵌套容易写错书写困难的问题函数括号嵌套阅读不直观的问题多个小函数的快速组装大函数函数临时变换使用方式函数中提前无法预知的一些异常的快速处理（不关注异常但是又不得不编写异常的问题，函数异常程序不退出继续执行，但是保留异常或者报警异常，不需要写一大堆的 try catch）

CHAPTER 10

后台程序执行过程的可视化 webview

动态和静态数据结构监控执行中的手工调试后台程序生成可视化界面给用户查看的

CHAPTER 11

jupyter 中调试编写难题

无法对动态数据展示使用 flask 等框架构建一个网页 webview 但是会被卡住无法执行其他代码执行一些循环会被卡住无法执行其他代码

CHAPTER 12

和操作系统交互等一些难点

提醒到移动设备问题手机报警，钉钉微信写文件读文件Ⓕ嗦语法系统文件目录和文件内容等增量更新监控大文件边读取边处理执行系统命令，获取系统命令输出数据问题

CHAPTER 13

流程设计和编排和计算难题

CHAPTER 14

实时持续产生数据的计算

实时持续来源数据的计算，未来数据的计算 window partision

多种输入输出模式 channel 源头（timer,redis,Kafka,request,tcpserver）单行道进出 p2p 双向通行 cs 一对多 topic 多对多网格 endpoint 死胡同点对流 Post to ,>>s 流对点 s>>endpoint 流对流 s1>>s2 去重复，flatten，去空值，过滤，map

stream 一切皆流，无物永驻