启动

爬取目录

node  D:\prj\spdJs\geneListPageFiles.js

爬取详细条目

//  node  D:\prj\spdJs\insertDetailinfo.js

下载视频查询与回调数据库接口

 node   D:\prj\spdJs\httpserver\_perlist.js

http://localhost:888

目录

[1. 爬虫技术说明 1](#_Toc21193)

[1.1. Webdriver （用于爬取目录和详细页数据 1](#_Toc10101)

[1.2. 浏览器插件tampermonkey （用于下载视频 1](#_Toc14297)

[2. 问题解决 1](#_Toc2854)

[2.1. webdriver调用Chrome不稳定，容易奔溃 1](#_Toc13511)

[2.2. 多线程问题 2](#_Toc31350)

# 爬虫技术说明

## Webdriver （用于爬取目录和详细页数据

因为pronhub做了很多跳转和js生成内容，使用普通的request库无法得到爬取目录和详细元数据源码。。所以使用此gui库驱动浏览器获取源码爬取。。

## 浏览器插件tampermonkey （用于下载视频

因为pornhub视频路径加密解密比较复杂，研究半天难度较大。。。所以写了个gui接口插件去调用下载它，简化操作

# 问题解决

## webdriver调用Chrome不稳定，容易奔溃

猜测可能插件加装太多，调用firefox驱动解决。。

## 多线程问题

试了process\_child模式，cluser模式均难以兼容webdriver ff。。。最后自己写小型线程池模式解决。。。