

# UDIAB——面向软件开发者的现代化搜索引擎

张曙 赵宏铎 钱正玭 熊吉思汗

2021年9月24日



- 01 项目主要背景
- 03 相关技术实现
- 05 项目时间安排

- 02 项目预期和优势
- 04 工作流程和测试部署





安装软件和编写代码过程中遇到一些细枝末节的问题。



使用百度搜索引擎发现一堆与本专业无关但关键词相似的网页



必须在众多网页中鉴别出相关的网页,解决问题效率低下。



安装软件和编写代码过程中遇到专业性较强的问题。



看完教程发现无法解决问题,并且前置各网站教程的复制粘贴率高。

搜到的:



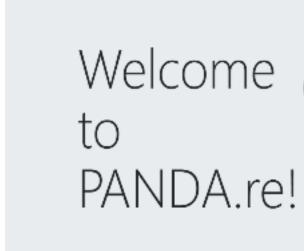




PANDA是一个二进制代码网络 安全领域的动态分析平台,它可 以通过录制和重放功能慢慢建立 起对数据的存储位置、运行的进 程、关键执行事件发生时间等的 理解。

Pandas是Python语言用于数据分 析的一个扩展程序库。 二进制代码安全方向的冷清加上 全民Python的热潮造成了搜 PANDA安装"却搜出全是关于 Pandas的使用教程。

### 希望得到的:



PANDA.re



PANDA is an open-source Platform for Architecture-Neutral Dynamic Analysis. It is built upon the QEMU whole system emulator, and so analyses have access to all code executing in the guest and all data. PANDA adds the ability to record and replay executions, enabling iterative, deep, whole system analyses. PANDA can be controlled from the command line, through our Python package, or even a Jupyter notebook.

### ▼ Whole-System Record and Replay

PANDA record whole system behavior. such that it can be subsequently analyzed iteratively and reproducibly.

### A Pre-Built VMs

We provide Linux QCOM images for many PANDA supported architectures: i386. x86\_64, arm, mips, and mipsel. These images are downloaded on

### Callbacks

PANDA provides over 40 callbacks at various points in our emulation code. Custom analyses can register functions to run at each of these locations.

panda平台 m.panda.tv xpanda平台怎么责张 panda交易所怎么样



Java

**建物设计与由用于**实

(2)1年11月23日 何自从Python进入3.99有智以Epondos的使用企構基础 智及 它的复数经常见于不被分价、原虫、金融分价以及科学计算中。 在 ·数据分析工具的技术或者,pendes作者普通 pendes中的可提化。

2019年(II.月21日 軟件管页: https://github.com/maybobanda 軟件介绍 PA. NDA是一个开票的平台网络中性致充分析,它是建立在CEMU全局的模

投票 所以分替获得所有代码的行案人的所有数据,PANDA提供

www.coderctu.com/sol/M172.... 〇 西蒙特爾

parataly国理学科V 编辑paratal软件 parataly领导 paratalization paratality.平台

平台契构中性动态分析 PANDA I 码次数件 - 码次网







和上述展示类似,针对这些日常工作学习上遇到的问题,我们分别对百度,必应和Google进行了一些词条(技术性文档和代码两个方面)的搜索测试,取前10个条目探索一下正确率。

搜索内容	百度	必 应	Google
路由器文件系统	6	8	1 D
panda亘 放 功 能	Ò	3	1
专业 作代 码 1	1	5	4
专业性代码2	D	6	3

傅卓驾走 格物明德 7



panda的录制和重放	0	百倍一下

○ 同页 回資訊 回視類 回向片 ②知道 □文库 明則那 △申兩 ①采购 更多 百亩法/核划到相关结果约85,000个▼投表工具

### 熊猫带着2.0.0版本更新,录制回放功能上线。

回顾应去的2020年,是欺英与高水并存供分和编艺商在《编书者也是在这样一个大环境下学育而生的。 例为们省在通过使用VR全需通路、图文等展示形式为多实现异地同算、现的互动、实

bbs2.720yun.com/article/id=1... 〇 百度快服

### 韩国女主播PandaGirls金宝拉直播录像回放-20170216 哔哩哔哩 (...



(22.37) Mile: 225.56 2020年8月8日 - YouTube

### 



(2.3) MH€ 46.38

2018年7月13日 - 能療16119Pands端2018年01月16日100651-105233自 播表學回改 是在依据使出的生活能清視數 于2018-07-13 21:48 01...

v youku.com/v\_showfd\_XMzcyM... 〇 百度快服

### 其他人还在提

pandakill录制后打架 pandakil大主真名 pandakill轉移同推荐 pandakill每pdd是模糊 pandakill制工时 pandakill编章数数 pandakill编纸 pandakill被拟人员

### 解猫直播韩国主播PandaGirl蔡宝美直播录像同放-20170107-生活-...



根據 D併长 93:23

2018年1月10日 - **城區里提幹区主播Pands C**前秦<u>安美国格灵修四教</u>-2017 0107 是在优配框出的主语高速模型于2018 01 10 14 30 16上线

v youku.com/v\_show(id\_XMTk0M... 〇 - 百良快期

### 輕猫 伊索婉 直播录像回放 20160428-娱乐-完整版视频在线观看-...



[22.3] MSH 119.45 2016年4月28日 - **於京伊京郊山南岩橋山**校20160420 www.lgW.com/w\_19rtz41f\_.hlml ○ - 百長株開

### 马洛卡VS比利亚雷亚尔 09月19日马洛卡VS比利亚雷亚尔全场。

2天前 00月19日乌姆卡VS比利亚雷亚尔**罗象 的程直接** IPS为森岩来乌姆卡VS比利亚雷亚尔全场**了** 使且放在纸刷着,还可以调着乌姆卡VS比利亚雷亚尔的历史交换**录使**回放。

www.xmymy.com/zugiu/luxiang/xt... 〇 白度快報

### 態苗 伊素婉 直播录像回放 20160428 高清在线观看 百度视频

2016年4月28日 **福廷** 伊京城 <u>国籍录像画教</u> 201604289555核力值 2017-42-28 微门推荐 搞美 职品 小品 海級 浙城 生活 科技 美食 宫乐 相关推荐05 23 白科树鱼味道枫不甜 1 3

百度機類 〇 百度快期

python也能做excel大神之一:鰕猫(pandas)的大用处

国内版 国际版 panda的 最後に重放 Q Q 阿瓦 即用 投票 サル 同意 総同

■ Repeatable Reverse Engineering with PANDA\_Chary() Blog ...

https://blog.csdn.net/qq\_32505207/article/details/104355205 •

**纳要** 动机 创新点 限

PANDA是一个开幕的结心分析平台。它建立在QFMU全系统的界限之上。网络分析可以访问符句正在执行 的代码并且可在客户也被扣提作所有数据。PANDA增加了记录和重数执行的能力。支持这代。汉人。完整 的系统分析。此外,重强日志文件是包含和可共享的。允许重复实验。PANDA利用QEMU对12种不同CPU 体系结构的支持,使得在LLVM-构中分析这些不同的指令重成为可能。PANDA分析是在一个简单的操作架 物中等等的。该架构包括一种在操作之间共享功能。增加分析代码重用和简化更多分析开发的机。

在blog csdn net上哥看更多信息

### 【十分钟Python知识点】pandas最详细数程 - 知乎。

https://zhuanian.zhihu.com/p/99889912 +

2009年16 - pandas 長有極的地方在于軍所強盛了很多也。它是一个統分也。軍所有很多其他包的功能。 按点視論。因为你只要要使用 pandas 既可以完成工作。 pandas 把当于 python 中 excel:它使用表(他 就是 dataframe)。能行数…

### [数件]视频录制 NVIDIA 即时重放:录制最近10分钟的视频。简书 https://www.janshu.com/p/e5ft7377d222 ▼

2019 4.7 · 即时里维的作用? "我"(一个管师用户),不需要从头到尼伞保费制能压内容,可能只需要保存一些自己应供助导程序的部分。比如INININI的针击数了一个8088,那么出现8088的第个边路就在问题过去的10分钟之内。这时候如果之前就已经开启了即时重放,那么就可以选择保存这最近10分钟的数数内容了

### Pandas数程 | 数据处理三板斧——map、apply、applymap ... https://zhuanian.zhinu.com/p/100064384 +

在日常的數據处理中,經常会对一个Ostaframe进行逐行。逐列到逐元素的操作。对应这些操作。Pendex中的map,applystapplymap可以解決绝大部分这样的數據处理需求。这篇文章或以底使附未匿解的方式。 为大家详细介绍一下位三个方法的实现更厚。相信媒先本文后,不给是小白还是Pandas的讲阶…

### A 前端实现录屏、跨页向录屏、截屏功能 - 熊猫程序员 - 博客园

https://www.cnblogs.com/panda-programmer/p/13342161.html + 2020-7-20 - mwet景其实现的原理是记录dem的变化。描绘的时候重新执行dom操作,所以不支持生成

2020-7-20 - mvet崇其其我的更是是定录dem的变化。接近的时候重新执行dom操作,所以不变持生战 mp4、只可以在web遭**需**放 hm/2 canvas截至实现的原理是主成一个canvas固有。因而的内容是操作器线 们实现好的。我们由受效的是将canvas最有保存或base642以,保存到

### 图吧小白教程 第二十一: 使用NVIDIA显示自带程序录制直播 ... https://post.smzdm.com/p/a809opi7 •

2019.11-11 · 各位部内分级模型十一块外。今天和第一个小白白粉样,希望各位喜欢、墨特500元的种样形。 的风声景。游戏开究直接 300元的郑朝藩·特灵游戏和郑美市AMD的方案,这个利且不妥 在游戏处理中。 我们经常会需要**要**刺自己游戏规度。记录自己经神的时刻,然而现在大

### 의 HackRF 无线门铃信号录制与重放 - OpenSourceSDR - 博客园

https://www.cnblogs.com/OpenSourceSDR/p/9602465.html = 2018-9-7 本文绘口类要求利用量改两个专家都可重观均有到和所到无线门较信号量改均效果。文绘环境: 提供平台—HackRF One, 天线 (40-800MHz), 无线门特 软件环境—Uhunth运行HackRF环境 系统 超率信息—Ubunth-14.04 (64-bit), 内核: 3.13.0-32-generic. 1) 打开效策、输入如下命令…

### ■ LoadRunner脚本调试与回放 QQ小惠的博客-CSDN博客

https://blog.csdn.net/thefirstray/article/details/78972999 -

2018-1-4 · 斯本泰利成功后,因改新本时。并不一定能够直接回放成功,可能是由于找不到的意文件,或 者缺少争四等亦作便因。一、Songett以进行版本资格。提用的原本原证可参考以下几点: 1、去约无关代码。 托姆凯索格加的语序来到可以去掉,比如"别等器性关节操"。 2、不平均是积明的"时间"输足下,thinkting





panda的录制和重放

Q 全部 日 图片 图 新闻 图 初版 ② 地图 : 更多

排程的 62,800 条结果 (用时 0.42 秒)

https://www.cnblogs.com > nullfun =

### 流量录制与回放技术实践- 田小波< - 博客园

2021年8月30日 — 文章导读本义主要介绍了流量录制与回放技术在区侧场景下的应用、通过调本稿文章、你将了解到开源的表别上具如何与内部系统集成、如何进行工况开发以... 铁少字词: penda | 必须包含: panda

https://iesterhome.com > topics +

### 录制线上流量做回归测试的正确打开方式 - TesterHome

2020年10月13日 — 人人节省构造则试数据,或者构造则试数据脚本的时间,提高效率。线上 量回放的限制是什么. 贝回放GET 请求. 因为其他请求的回放,会对用户数据...

缺少字词: penda | 必须包含: penda

https://segmentfault.com/ ... v

### 性能工具之常见流量复制工具 - 思否

2021年6月11日 — Sharingan(中文名:写轮眼)是一个总于golang 的流量录取回放工具,最终上真实选求流量进行回放测试,适合项目重构、回归测试等、流流升源; 地址: .... 缺少字词: panda | 必须包含: panda

http://kw.hscode.net > Allterns > kw=Panda USB无线w... \*

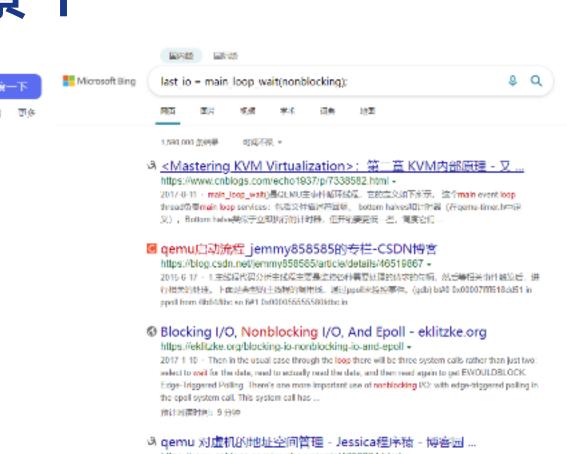
### 创建智慧型物流生态链 - 通关网

更多>>, 音询结果: Panda USB无线网卡液发发射器 数据更新: 2021年09月06日 ... 其他使用









https://www.cnblogs.com/wuchanming/p/4732604.html -

2015-6-22 - 可以看到整个线程除了进入kvm没有加铁,其他时候都会加铁。也就是没vcsu线程里处理in事 什的时候是交持有这把锁的。 再看看这把锁在gemu里的应用: 在os\_host\_main\_loop\_wait甲有这把锁的

### ✓ Non-blocking console input C++ - Stack Overflow

https://stackoverflow.com/questions/6171132 +

2021-3-22 - I'm looking for a (multiplatform) way to do non-blocking console input for my C++ program. so I can handle user commands while the program continually runs. The program will also be outputting information at the same time.

I would do this by creating separate a thread which calls normal blocking IO functions and pass it a callback function which it... 現住答案 ・ 申 13

Example using C++11: Maclude clostroams Ainclude «future» Ainclude <thread> #Include <chrono> static std:string getAnswe...

Eve done this on QNX4.5 that doesn't support threads or Boost by using select. You basically pass select STDIN as the file descript...

awyne wait / non blocking wait in python. C# non-blocking socket without while(true) loop. 2013-5-22 2009-11-3

### 查得更多纳里

### 以 IO通信模型(二)同步非阻塞模式NIO(NonBlocking IO ...

https://www.cnblogs.com/niumoo/p/9847154.html •

2018-10-25 · 同步北阳整模式 (NonBlocking IO) . 在北阳整模式中,发出Socket的 accept () 和 read () 接作时,如果内核中的数据还没有准备好。那么它并不会能塞用户进程。而是立刻返回一个信息。 他然是 资法程发起一个resid操作后,并不需要一直租拿等待,而是马上被得到了一个...





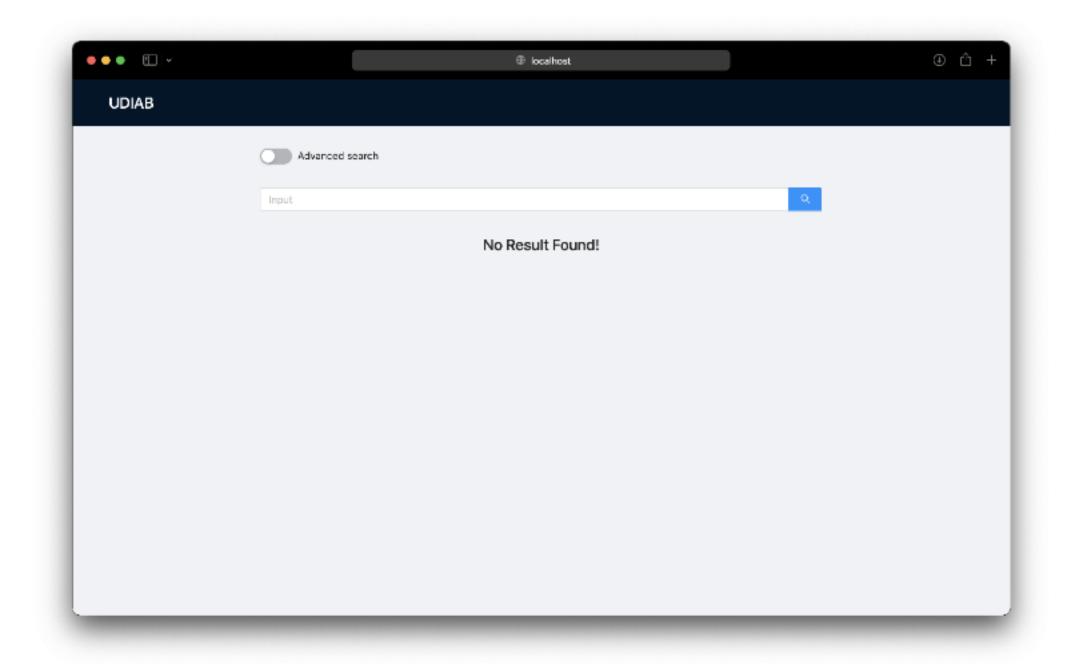
2020年9月23日 — 非组基FD (nonblocking FD): 通过进程反复调用 O 函数,在数据拷贝过程 profile getolock(); #endifilast to = main loop wait(nonblocking); ...

傅早驾走 格物明德



# 项目预期效果





实现对上面现有搜索引擎的优化,着重解决专业性问题搜索的不足。包括用户针对技术性文档的搜索需求和代码搜索方面的需求。

傅卓驾志 格物明德

# 项目预期优势



## 针对性搜索

我们将用户聚焦为软件开发人员,从软件开发人员评价较高的活跃的专业平台、论坛中获得检索文档库,尽可能使检索的正确率较其他搜索引擎有显著的提高。同时,我们利用更加高效、现代的技术和轻量级的模块尽可能使信息检索的等待时间较其他搜索引擎有提高。

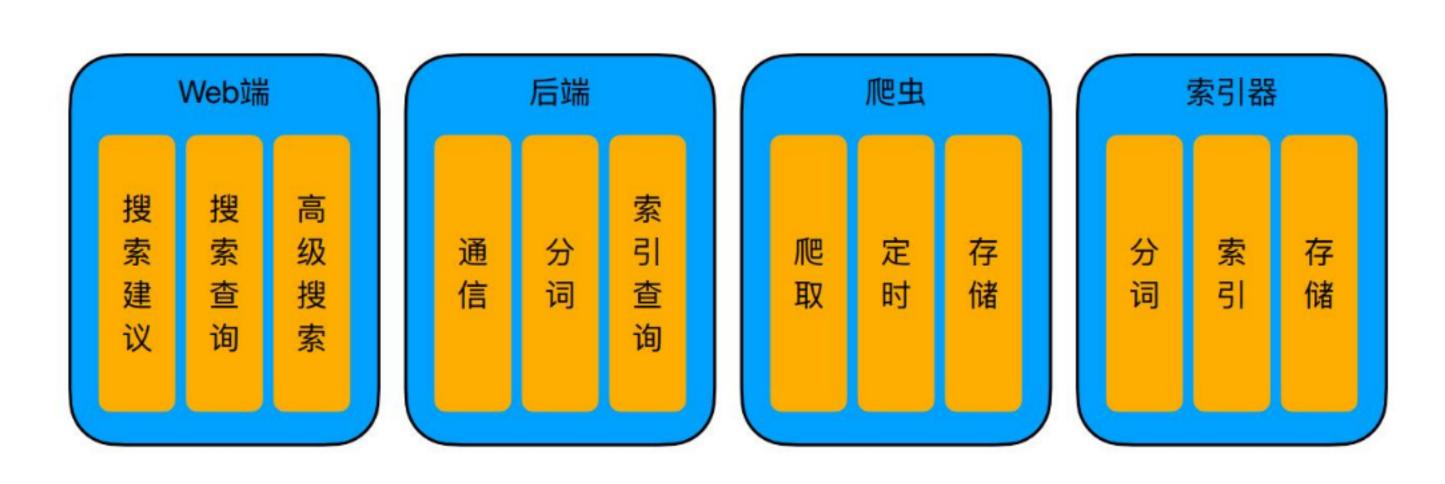
# 代码搜索

UDIAB鼓励用户通过代码进行搜索,为代码信息设计特殊的分词方式,使通过 代码搜索变得更加智能

傅早驾走 格物明德

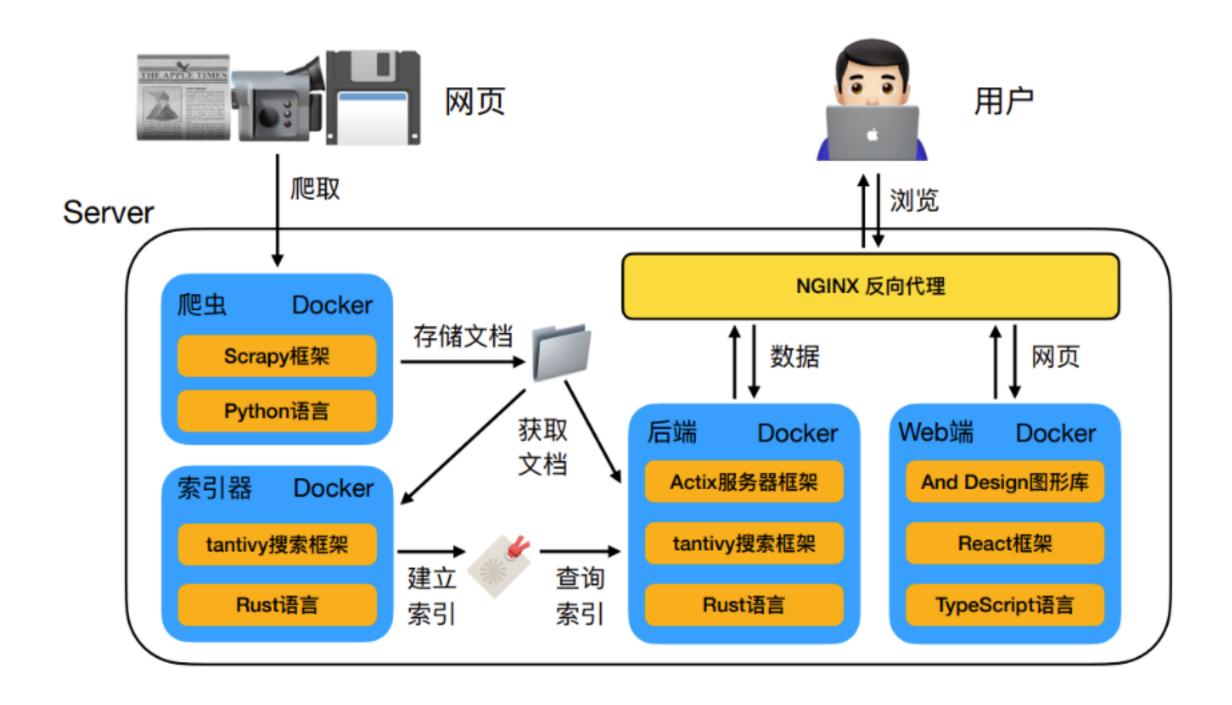


# 项目功能架构





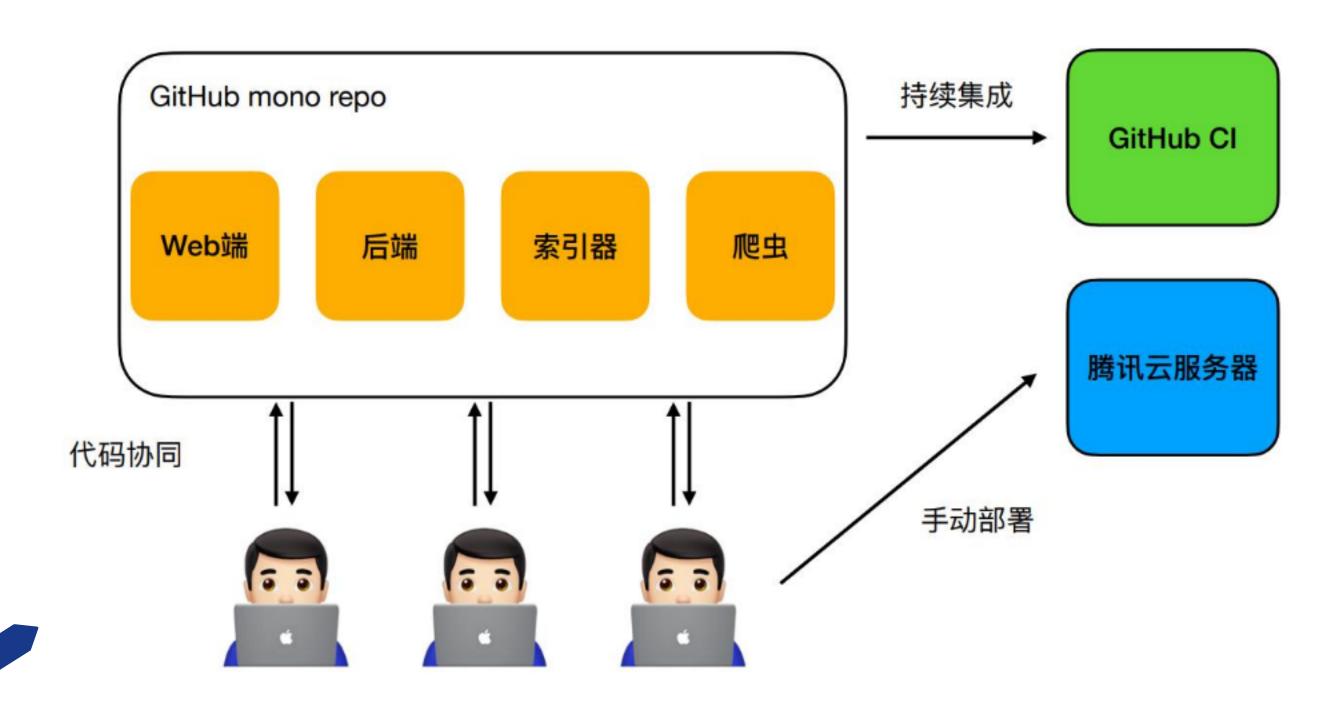
# 技术选型







# 工作流程和测试部署





# 人员分工





张曙

项目的Web端 和后端的开发等



赵宏铎

项目的文档编 写和编写索引器 代码等



钱正玭

编写爬虫代码, 实现定期数据爬 取功能等



熊吉思汗

进行项目测试, 完善代码等

# WEB端及后端



组件	进度安排		
组件	10.1 ~ 10.17	10.17 ~ 11.17	11.17 ~ 12.17
WEB端与 后端	1. 设计实现WEB端搜索查询模块并进行单元测试。 2. 设计WEB端与后端交互的接口。	1. 按照通信模块->搜索建议模块->分词模块->搜索查询模块->高级搜索模块的顺序依次进行设计实现并依次完成对应单元测试。	试。

**得早然志 格物明德** 

# 爬虫组件



组件	进度安排		
组件	10.1 ~ 10.17	10.17 ~ 11.17	11.17 ~ 12.17
爬虫	2. 对拟定的平台、论坛进行爬	块。	<ol> <li>设计实现定时模块。</li> <li>对整个爬虫组件进行功能测试。</li> <li>完善爬虫组件。</li> </ol>

**得学驾志 格物明德** 

# 索引器组件



组件	进度安排		
组件	10.1 ~ 10.17	10.17 ~ 12.1	12.1 ~ 12.17
索引器	1. 学习语言、Tanvity搜索框架和Jieba分词系统的相关知识。 2. 设计针对代码的分词方式。	1. 按照分词模块->索引模块-> 存储模块的顺序依次进行设计 实现并依次完成对应单元测试 。	0

**得学笃志 格物明德** 22

#