



UDIAB——面向软件开发者的 现代化搜索引擎

张曙 赵宏铎 钱正玳 熊吉思汗

2021年9月24日

目录

01

项目主要背景

02

项目预期和优势

03

相关技术实现

04

工作流程和测试部署

05

项目时间安排



Part 01

项目主要背景

项目主要背景-1

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。

项目主要背景-2

安装软件和编写代码过程中遇到专业性较强的问题。



使用百度等搜索引擎发现以CSDN为首的一大堆技术网站



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。

项目主要背景-2

搜到的:

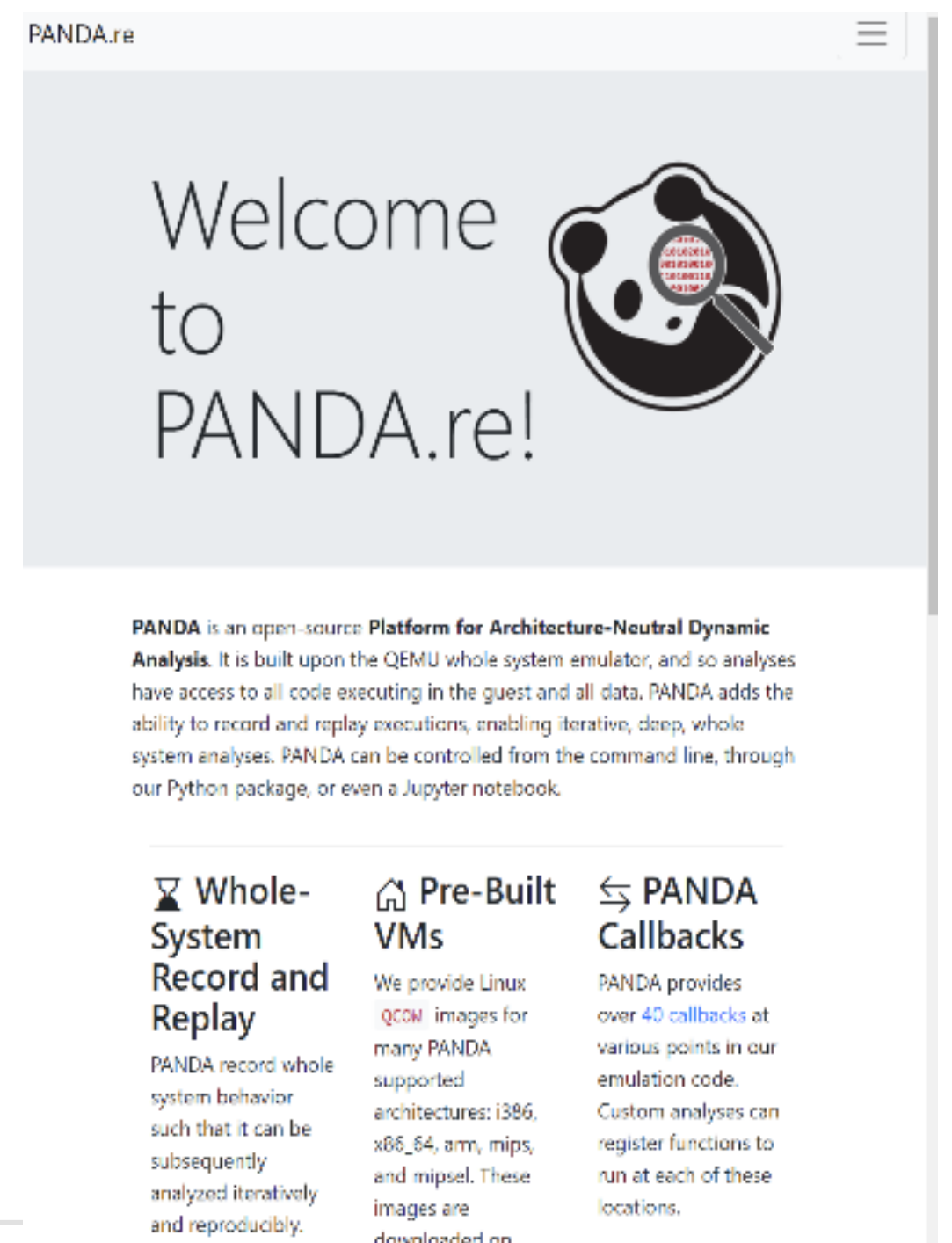


PANDA是一个二进制代码网络安全领域的动态分析平台，它可以通过录制和重放功能慢慢建立起对数据的存储位置、运行的进程、关键执行事件发生时间等的理解。

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:



项目主要背景-2

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	G o o g l e
路由器文件系统	6	8	10
p a n d a 巨 放 功 能	0	3	1
专 业 栏 代 码 1	1	5	4
专业性代码2	0	6	3

项目主要背景-2



项目主要背景-1

 last_io = main_loop_wait(nonblocking);

百度搜索

百度为您找到相关结果约2,200,000个

百度搜索

qemu启动流程_jemmy858585的专栏-CSDN博客

2015年6月17日 last_io = main_loop_wait(nonblocking); while (!main_loop_should_exit()) { main_loop_should_exit函数分析,判断是否有关闭,暂停的请求。 static pool_ma...

来自 CSDN技术社区 0 点赞快评

【QEMU-KVM代码分析之三】IO thread源码浅析之main_loop_L...

2015年8月11日 int main_loop_wait(int nonblocking) { int ret; uint32_t timeout = UINT32_MAX; if (non_blocking) { timeout = 0; } /* poll any events 1/2 XXXX: separate dev...

来自 CSDN技术社区 0 点赞快评

qemu 对虚拟机的地址空间管理 - Jessica程序猿 - 博客园

2015年8月15日 main_loop () { do { last_io = main_loop_wait(nonblocking); ... } while (!main_loop_should_exit()); } main_loop_wait () { ret = os_host_main_loop...

博客园 0 点赞快评

其他人还在找

blocking 什么意思啊 nonblocking non pollring nonblocking pool loop模型 为什么用blocking blocking fail antblocking blocking voltage blocking alias

5. <mastering kvm virtualization>:第二章 kvm内部原理(1)

2017年8月11日 static void main_loop(void) { bool nonblocking; int last_io = 0; ... do { nonblocking = kvm_enabled() && !kvm_enabled(); && last_io > 0; ... last_io

博客园 0 点赞快评

qemu aio笔记-sak0-ChinaUnix博客

2015年7月11日 req->ret = ret; /* 从IO设备驱动event_notifier_poll(&pool->notifier);主线程main_loop () { last_io = main_loop_wait(nonblocking); ret = os_host_main_loop...

m.blog.chinaunix.net/uid-20718... 0 点赞快评

最全的剖析QEMU原理的文章3-xiong8937-ChinaUnix博客

2015年4月22日 static void main_loop(void) { bool nonblocking; int last_io = 0; do { nonblocking = !kvm_enabled(); && last_io > 0; last_io = main_loop_wait(nonblocking);

blog.chinaunix.net/uid-2285422... 0 点赞快评

我开的服务器崩溃了!!求大佬分析一下原因【minecraft吧】...

回复数 7 发帖时间 2021年01月20日

1楼 ...FileNotFoundException in server tick loop.java...


1楼 小弟想请问下, 开这个线程, 在端人不会有问题...

百度贴吧 0 点赞快评

...Stopping and Hashing operations are blocking MainL...

2021年1月18日 Also - I look for taking the time to double check what was blocking the engine for rml without...

来自 CSDN技术社区 0 点赞快评

 last_io = main_loop_wait(nonblocking);

Microsoft Bing

1,540,000 条结果 时间不限

<Mastering KVM Virtualization>: 第二章 KVM内部原理 - 又 ...

https://www.cnblogs.com/echo1937/p/7338582.html

2017-0-11 · main_loop_wait是QEMU主事件循环函数。它的定义如下所示。这个main event loop thread负责main loop services: 检查文件描述符就绪、bottom halves和计时器 (在qemu-timer.h中定义)。Bottom halves类似于立即执行的计时器,但开始要更低一些,调度它们...

qemu启动流程_jemmy858585的专栏-CSDN博客

https://blog.csdn.net/jemmy858585/article/details/46519867

2015-6-17 · 1.主线程代码分析主线程主要负责处理各种需要阻塞的请求的回调,然后等待相关事件发生,进行相关的处理。下面是典型的主线程的伪代码。通过poll来监听事件。(qemu) bsp0 0x0000700518a51 in epoll from 0x0000700518a51 in 0x0000555555801d8c in

Blocking I/O, Nonblocking I/O, And Epoll - eklitzke.org

https://eklitzke.org/blocking-io-nonblocking-io-and-epoll

2017-1-10 · Then in the usual case through the loop there will be three system calls rather than just two: select to wait for the data, read to actually read the data, and then read again to get EWOULDBLOCK. Edge-triggered Polling. There's one more important use of nonblocking IO: with edge-triggered polling in the epoll system call. This system call has ...

预计阅读时间: 9 分钟

qemu 对虚拟机的地址空间管理 - Jessica程序猿 - 博客园 ...

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

2015-6-22 · 可以看到整个线程除了进入kvm没有加锁,其他时候都会加锁。也就是说vcpu线程里处理v事件的时候是保持锁加锁的。再查看该函数在qemu里的应用: 在os_host_main_loop_wait里有加锁的...

Non-blocking console input C++ - Stack Overflow

https://stackoverflow.com/questions/5171132

2012-3-22 · 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...	Example using C++11: #include <iostream> #include <iatomic> #include <atomic> #include <chrono> static std::string getAnswer...	I've done this on QNX6.5 that doesn't support threads or Boost by using select. You basically pass select STDIN as the file descript...
最佳答案 · 13	18	5

async wait / non blocking wait in python 2013-5-22

C# non-blocking socket without while(true) loop 2009-11-3

查看更多结果

 last_io = main_loop_wait(nonblocking);

Google

找到约 1,320 条结果 (用时 0.34 秒)

https://www.geck-share.com/detail/

qemu启动流程 - 极客分享

2015年6月17日 — last_io = main_loop_wait(nonblocking); while (!main_loop_should_exit()) main_loop_should_exit函数分析,判断是否有关闭,暂停的请求。

http://blog.chinaunix.net/uid-22954220... 转为简体网页

最全的剖析QEMU原理的文章1 - ChinaUnix博客

2015年4月22日 — void main_loop_wait(int nonblocking) { rlds = -1; FD_ZERO(&rlds); ... last_io > 0; last_io = main_loop_wait(nonblocking); } while ...

http://blog.chinaunix.net/uid-29141031-id-5090787

qemu aio笔记 - ChinaUnix博客

2015年6月21日 — event_notifier_set(&pool->notifier); 主线程 main_loop () { last_io = main_loop_wait(nonblocking); ret = os_host_main_loop_wait(timeout_us);

http://www.resery.top/ 2020/06/23

QEMU KVM Note II | Resery's_blogs

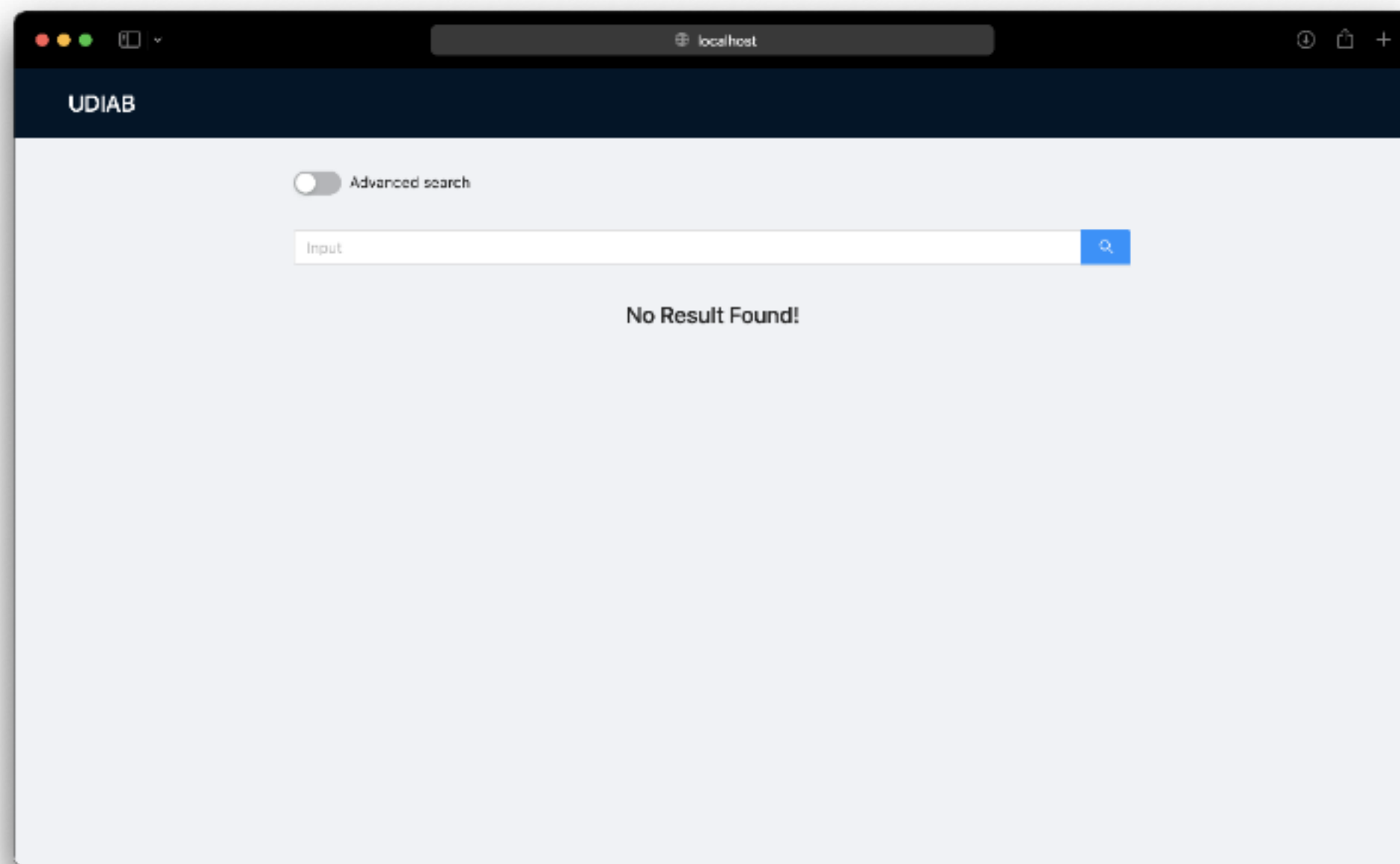
2020年9月23日 — 非阻塞I/O (nonblocking I/O): 通过并反复调用O 函数,在数据读写过程 profile_getclock(); #endif last_io = main_loop_wait(nonblocking); ...



Part 02

项目预期和优势

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。

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

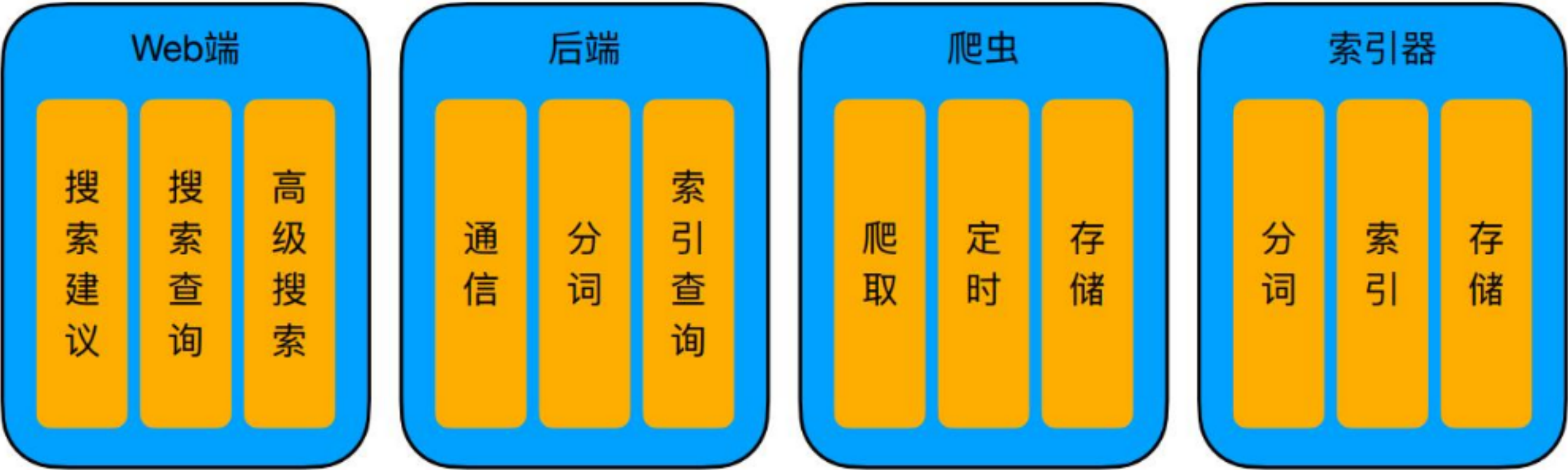
UDIAB鼓励用户通过代码进行搜索，为代码信息设计特殊的分词方式，使通过代码搜索变得更加智能



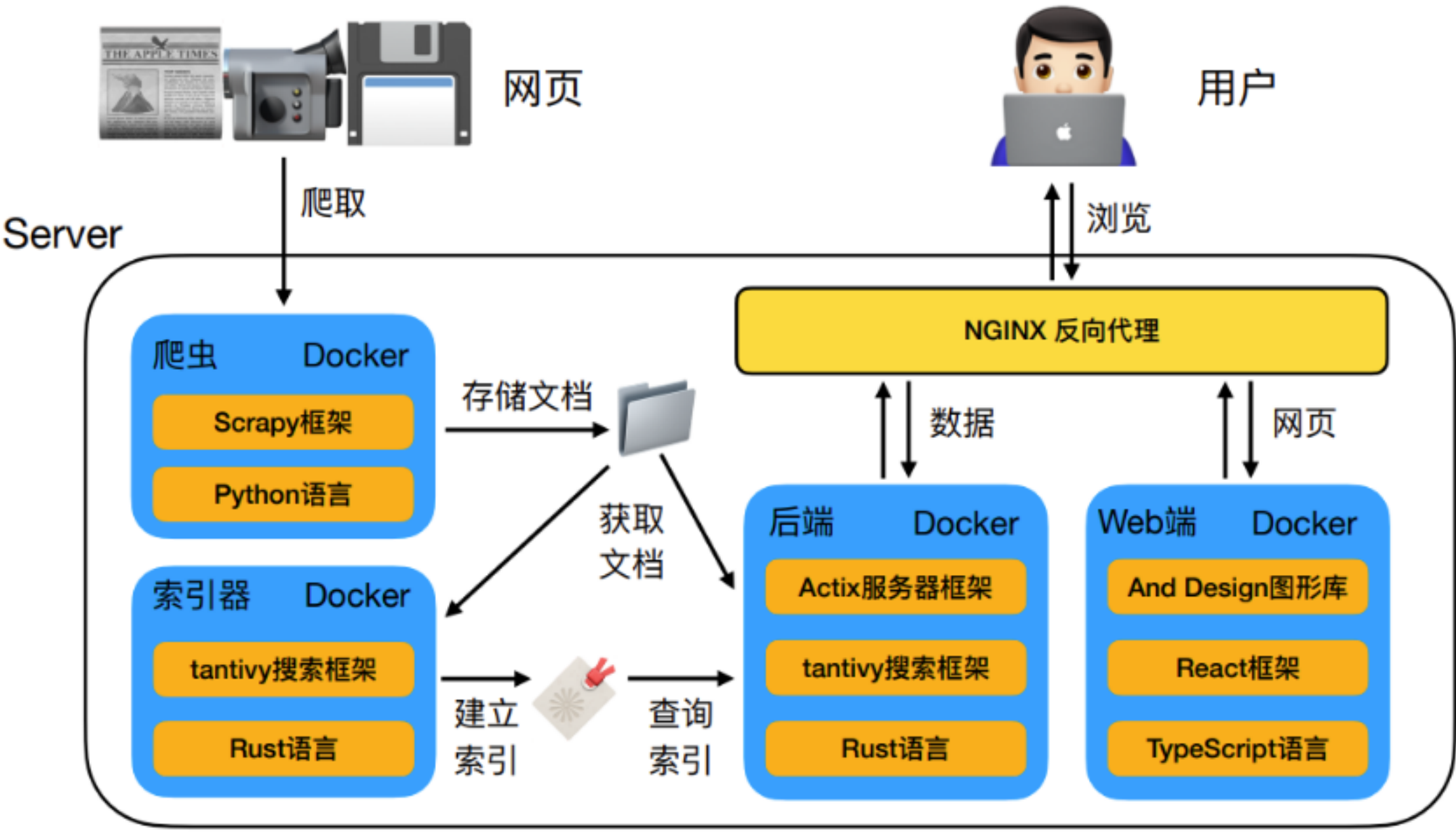
Part 03

相关技术实现

项目功能架构



技术选型

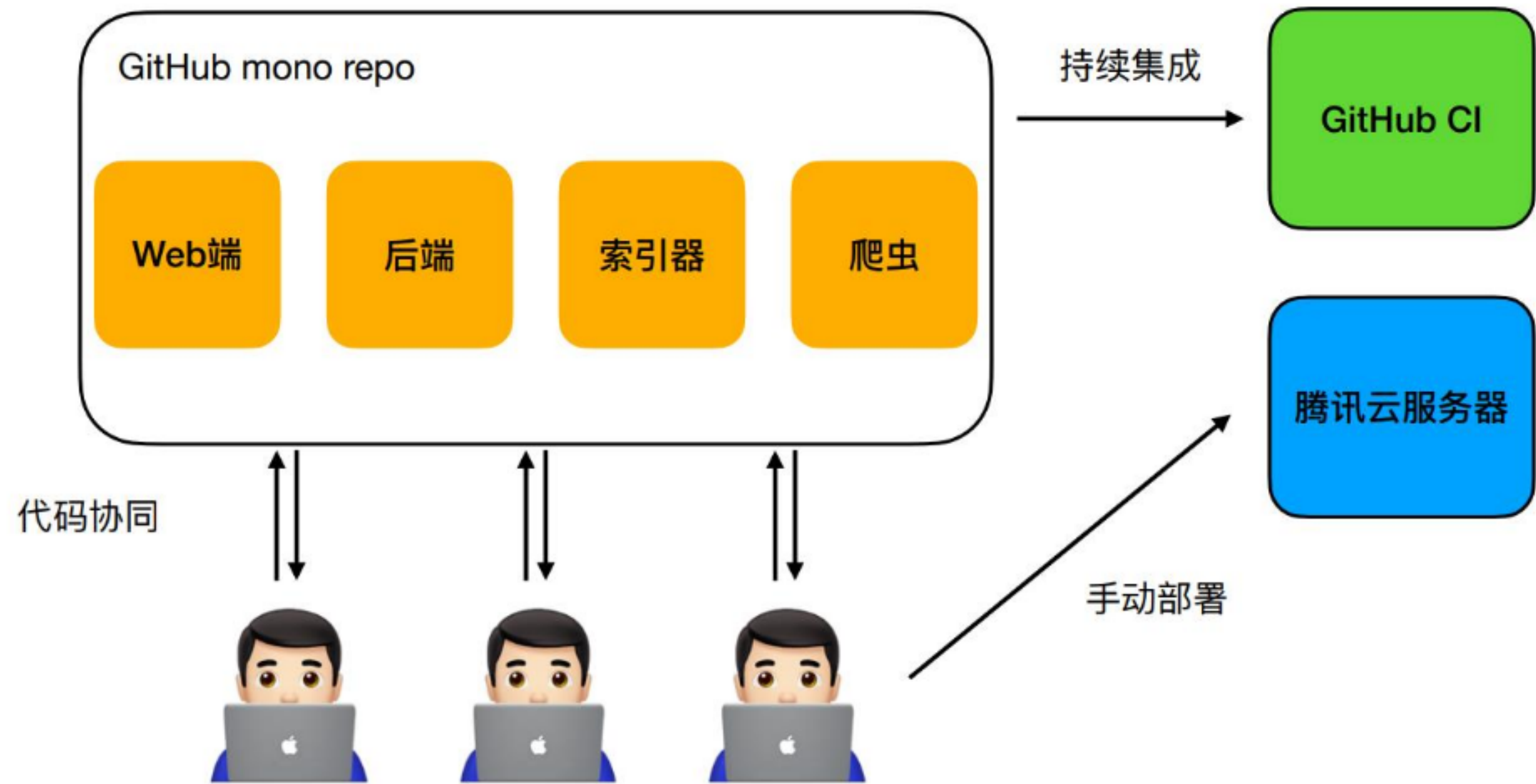




Part 04

工作流程和测试部署

工作流程和测试部署





Part 05

时间安排

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



钱正玼

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

组件	进度安排		
	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。	1. 对WEB端和后端进行功能测试。 2. 进行UDIAB集成测试。

组件	进度安排		
	10.1 ~ 10.17	10.17 ~ 11.17	11.17 ~ 12.17
爬虫	1. 学习爬虫框架。 2. 对拟定的平台、论坛进行爬取测试并确定爬取文档的平台、论坛。 3. 确定存储数据的格式。	1. 设计实现爬取模块和存储模块。 2. 对爬取模块和存储模块进行单元测试。	1. 设计实现定时模块。 2. 对整个爬虫组件进行功能测试。 3. 完善爬虫组件。

组件	进度安排		
	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。 。	1. 对索引器进行整体功能测试。 2. 进行UDIAB集成测试。

Q 问答 & A