

X站之痛-如何确保内容安全?

KUAIYE 快页

汇报人：吴青松

汇报时间：2022.9.8

目录

背景介绍

解决方案

案例分享





COVER NEWS

四川互联网联合辟谣平台

中共四川省委网信办主管
四川省网络治理研究中心主办
封面新闻承办

地震发生时男子上厕所被困男子！
就跑出来了！

注意！

关于四川泸定
6.8级地震

这些消息
都是谣言！

流言粉碎机

9月5日 12:52, 四川省泸定县发生了6.8级地震。地震发生后, 社交平台上出现了不少与地震相关的不实信息。四川互联网联合辟谣平台对热传涉地震谣言进行了梳理。提醒大家: 不要轻信未经证实的消息!

谣言: 地震时有男子高楼跳伞逃生?

真相



地震发生后, 一张男子从高层建筑上跳伞逃生的照片在微信传播, 有网友配文称: “跳伞下楼的, 是双流东立国际小区”。成都市双流区互联网新闻信息中心辟谣: 经查, 这是一张老图, 早在2013年时即有传播。有网友指出图片来源为“重庆某家滑翔伞广告”。



今年4月以来，中央网信办会同相关部门开展为期三个月的“清朗·整治网络直播、短视频领域乱象”专项行动，集中整治“色、丑、怪、假、俗、赌”等违法违规内容呈现乱象，从严整治功能失范、“网红乱象”、打赏失度、违规营利、恶意营销等突出问题。

专项行动期间，各直播、短视频平台积极开展自查自纠，针对问题全面深入整改，依法依规主动处置处罚违规行为。抖音、淘宝、微信视频号等16家重点直播、短视频平台共处置违规直播间**56.3**万个，清理违规短视频**235.1**万条，关闭违规用户账号**12**万个，处置处罚违规主播、短视频账号**21.86**万个，推动网络直播、短视频行业信息内容呈现面貌实现初步改观。

整治期间，各级网信部门根据网民举报、督导检查、网上巡查发现的线索，依法查处存在信息安全管理漏洞、问题整改不及时不彻底的网站平台**134**件次，予以行政处罚**1200**余万元。

举报入口

您可选相应入口，举报网上有害信息



政治类



暴恐类



诈骗类



色情类



低俗类



赌博类



侵权类



谣言类



其他类





为什么违规内容总是出现?

①黑产非法广告

②想成名的人

③自身价值观问题

④安全意识疏忽

⑤其他居心叵测的情况

月入过万的内容审核，是进入大厂的捷径吗？

播报文章



黑马程序员

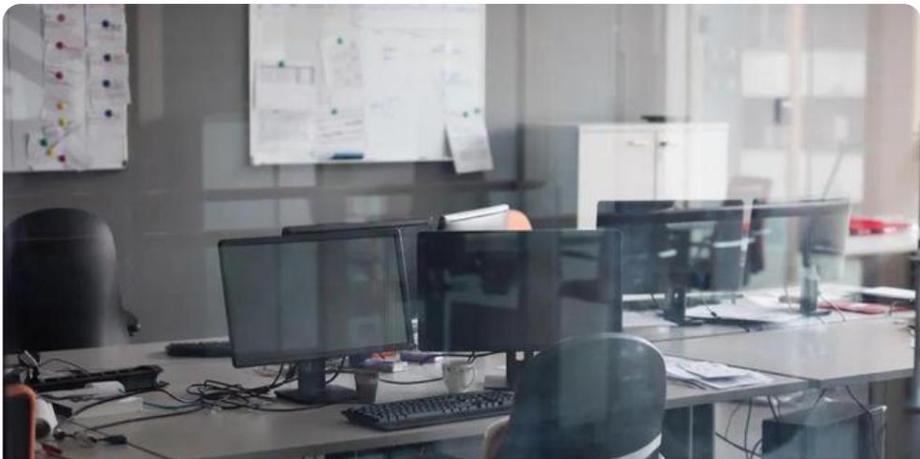
2022-05-19 15:56 北京 | 黑马程序员官方帐号

关注

内容审核员这个岗位，对学历要求不高，工作经验不多的年轻人来说。是接触互联网和大厂的机会。

据悉，某站有2413名员工属于内容审核岗位，字节审核员超过2万人。随着清明，以及网络平台管理规范的推出。各个企业扩招审核岗位的规模。

时至今日，腾讯、百度、阿里、快手等互联网公司都建立了内容审核团队。据报道，b站目前增加了1000个审核员岗位。



B站有员工过年因加班猝死？回应：事发前一周内不存在加班情况

2022-02-07 15:27:29 来源: 红星新闻

举报



8391

分享至



2月7日，有多名网友向职场博主@王落北爆料称，B站（即哔哩哔哩，09626.HK）武汉AI审核组组长管某（花名“暮色木心”）因过年期间被要求加班，在晚上九点工作到早上九点的工作强度下，于2月5日凌晨脑出血猝死。



国内知名开发网站联合开发网因涉政违法被罚50万 面临倒闭

大事件 1年前 0 1.5k

圈内比较知名的编程技术网站：**联合开发网** - pudn.com最近踩雷了，原因是**网站内的帐户发布涉政违法信息**，多个**帐户名称违法违规**，性质恶劣，情节严重。

我办于2021年4月21日对你主办的**联合开发网**用户账号发布涉政违法信息的行为进行立案调查，现查明**联合开发网**用户账号发布涉政违法信息，多个用户账号名称违法违规，性质恶劣，情节严重。

以上事实有电子数据取证、调查笔录等证据事实。

现根据《网络安全法》第四十七条、第六十八条第一款之规定，决定给予你限期整改**联合开发网**、暂停新用户注册，罚款50万元的处罚，限你于2021年6月25日前整改到位，整改期间，全面排查清理违法账号和信息，暂停新用户注册。

你应在收到本决定书之日起十五日内，持本决定书和《湖南省非税收入一般缴款书》缴纳罚款（银行：长沙银行



The screenshot shows the pudn.com website interface. At the top, there's a navigation bar with '首页', '资讯', '分类', and '金币商城'. A search bar is present with the text '问题没解决? 不妨搜一搜'. On the right, there are links for '加入VIP 赠3月' and '传资源得金币|悦礼品' along with a '登录' button.

The main content area is titled '热门分类' and lists various categories: 源码/资料, Windows编程, Linux/Unix编程, WEB开发, 移动互联网开发, 数值算法/人工智能, 区块链开发, 网络编程, 系统/网络安全, 通讯编程, 游戏, 多媒体, 嵌入式/单片机/硬件编程, 图形图像, 物联网, Java编程, 行业应用, 数据库系统, 虚拟/增强现实-VR/AR, 压缩解压, and 更多分类.

Below this is a '最新上传' section with a 'New' badge. It lists several uploads:

- [源码/资料] WordPress开发的智能SEO优化插件, 与众多WordPress的SEO插件不一样的, 更加简单易用. 378.4KB, PHP, 0次下载, 2022-09-06 21:40:38.
- [电子书籍] Dive Into Algorithms. 4.7MB, Python, 0次下载, 2022-09-06 17:51:05.
- [其他] 阅读器pdf. 3MB, Others, 0次下载, 2022-09-06 17:25:03.
- [源码/资料] 工控通讯协议, 以及解析. 726KB, C#, 0次下载, 2022-09-06 17:12:55.
- [源码/资料] SolidWorks 2020 - Step-By-Step Guide. 46.3MB, Others, 0次下载, 2022-09-06 16:45:43.
- [源码/资料] 卡尔曼滤波进行MPU6050姿态解算. 5.2KB, C/C++, 0次下载, 2022-09-06 15:51:26.

On the right sidebar, there's a '热门资源' section with a '周 | 月' selector. It lists 10 items:

- dianjizhuansuceliangxitong.zip
- 数字滤波器的 MATLAB 与 FPGA 实现...
- 分数阶混沌系统: 分数阶混沌系统的...
- 奥本海姆《信号与系统》中文版-MIT...
- UBLOX GPS资料包V5.2.rar
- 通信信号盲均衡算法
- WSNM_RPCA_p.zip
- 高精度捷联惯性导航系统Matlab工具...
- bianjiecheng.rar
- 遗传算法优化出入库路径.zip

At the bottom right, there's a promotional banner for '分享优质内容, 连接铁杆粉丝' with a '去上传' button and the text '与更多同好在开发者领域尽情翱翔'.

效率难题

面对巨大的数据流，人工审核工作遇到种种难题。

一是工作量巨大，需要消耗大量的人员和物资成本以及巨大的时间成本。

二是在面对复杂的视频内容的时候人工容易出现误判，并且人工无法实现系统一样对敏感素材库的记忆和管理。

还有在面对不断出现新的敏感素材的时候无法进行及时更新等一系列的问题，这些问题都需要系统工具去帮助其解决。

技术难题

在视频内容剧增的情况下，原有的文字识别技术和传统的文件级比对技术以及无法满足当下的需要文字、视频、图片等全方位检测技术需求。

视频中画面不断的变换，各种干扰的情况都会出现，例如裁剪、涂鸦等，传统的技术面对干扰无法准确识别，需要更抗干扰的识别技术去实现。

在不断出现相似的不良信息的时候需要能够识别相同类型素材的识别能力。

目录

背景介绍

解决方案

案例分享



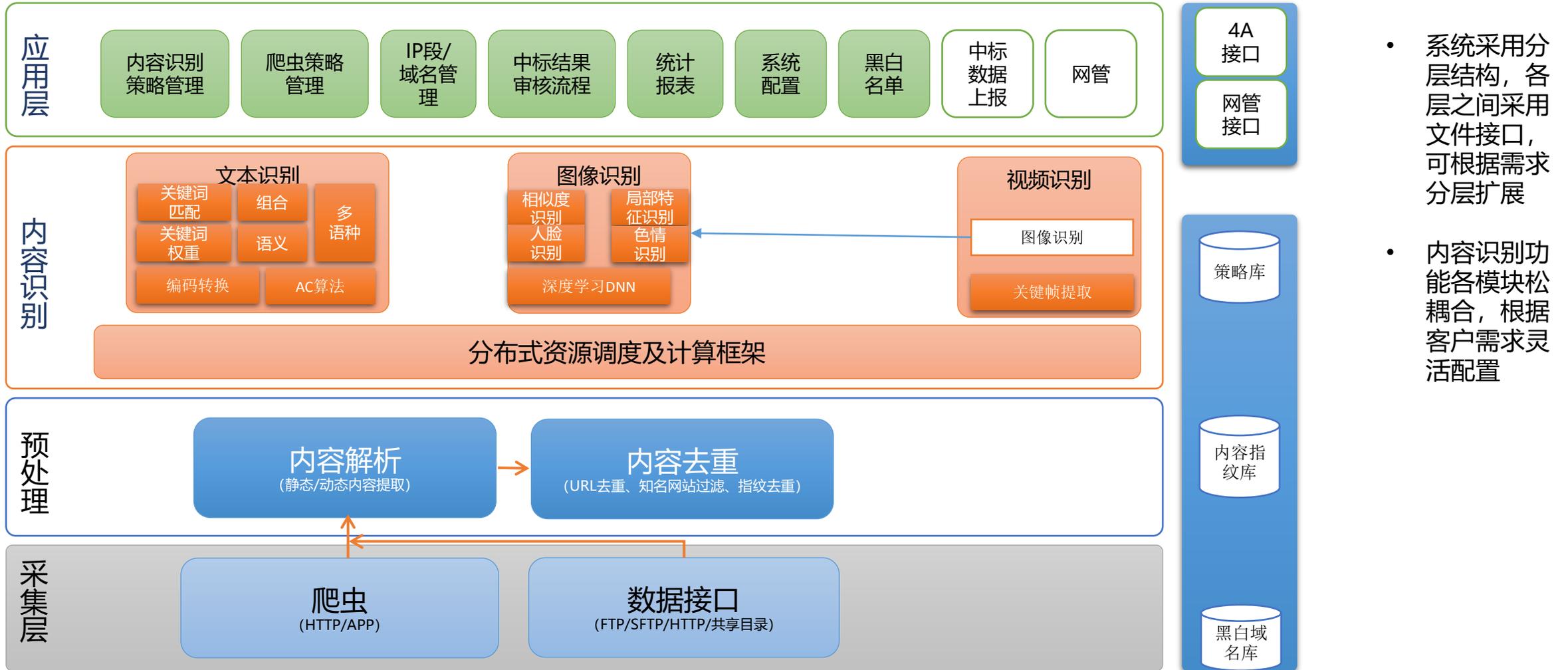
做内容安全工作的基本原则是：在保障内容安全的基础上尽量降低成本、减少影响。

智能审核流程图



机器审核

人工审核



●灵活配置

IP段、域名、URL可配置

爬取周期、爬取深度、爬取广度、调度优先级可配置

链接提取规则、结构化数据规则可配置

●自动化

7X24小时无人值守运行

异常站点自动告警

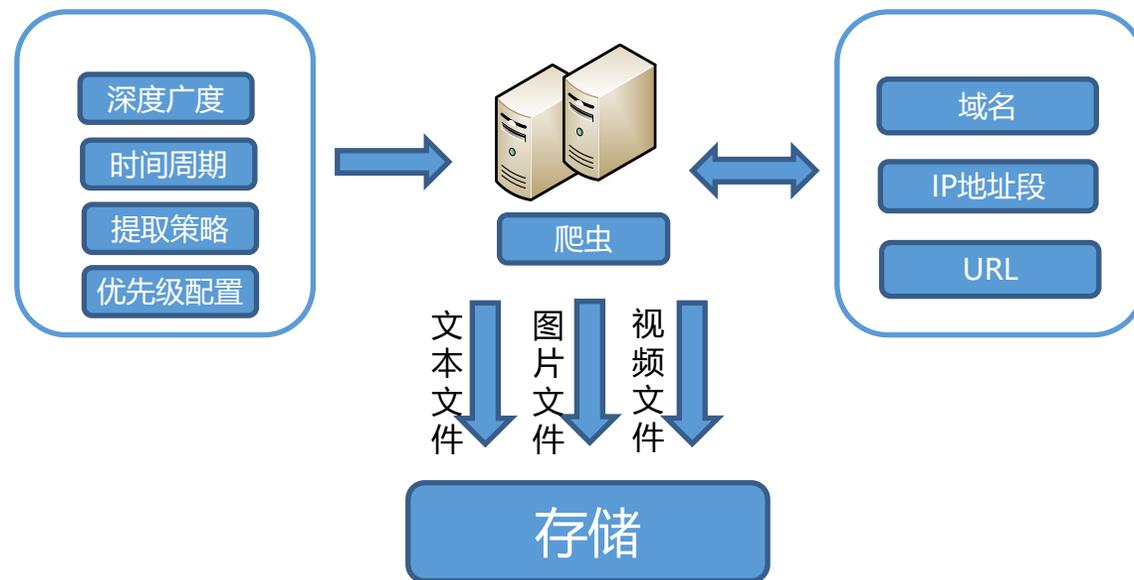
子链接、结构化数据自动化提取

●反反爬虫策略

不同User Agent模拟

Session/Cookie支持

代理状态实时监测和切换



爬虫架构

- **基础爬取模块**：通过模拟发送Http请求实现对待爬取URL资源流的获取并可直接存储到本地进行分析利用。
- **链接提取模块**：通过定制规则配置（也支持无规则配置的全站式连接提取）对爬取到的资源文件进行自动解析，并自动提取可用链接，从而实现对站点的深度爬取。
- **结构化数据提取模块**：通过定制规则进行爬取内容的自动结构化数据提取。

- **代理管理模块**：出于对部分站点的反爬虫策略的应对，爬虫系统支持代理模式，能迅速在对端IP封堵策略下及时切换代理从而实现不间断爬取。
- **视频地址自动解析模块**：对于大部分的视频站点，爬虫系统支持对视频播放页面的隐藏视频地址进行自动解析提取，从而实现在线视频的爬取。
- **分布式管理模块**：可实现爬虫主机的热加载和热卸载，自动完成爬取任务的均衡分配和合理调度。

● 必要性

- 互联网长尾效应：80%的访问请求20%的内容

● 去重方法

- URL级别去重
 - MD5比对法
 - Hash表配合URL压缩法
 - Bloom Filter去重
- 文件唯一编码级别的去重
 - 一般通过文件MD5进行相同文件的去重
- 文件指纹级别的去重
 - 使用我公司自研的指纹提取和比对技术进行文件相似性比对去重

● 不同采集方式的去重

- DPI采集方式进行DPI设备一次去重和平台二次去重
- 爬虫/文件采集方式进行平台一次去重

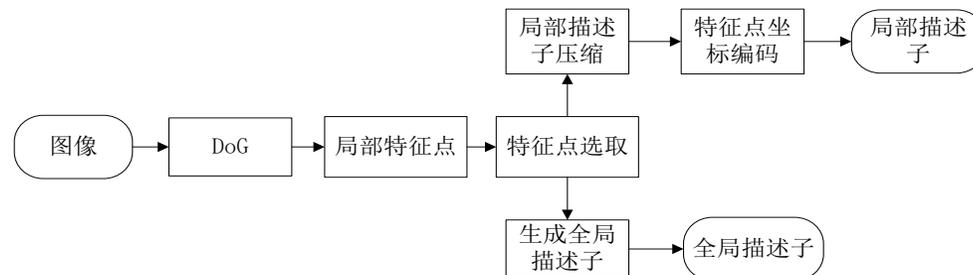
降低识别数据量

50%~90%

去重效果

● 图像及视频指纹

指纹即提取图像或视频的特征。我们的图像指纹由全局描述子和局部描述子两部分组成，全局描述子用于建立数据库索引，系统利用全局描述子快速的从数据库中筛选出可能相似的图像，然后利用局部描述子进一步计算检索图像与筛选出来的候选图像的相似度，然后根据相似度从高到低将结果返回。下图为图像指纹的提取过程：



智能语义分析技术主要包括：自动分词、关键词匹配、权重计算、模糊匹配

识别准确率>95%

关键词匹配

- 1) 基于关键词库，对文本内容进行全文检索，实现关键词精确匹配
- 2) 单机支持关键词数量>5000个，关键词长度不限

关键词组合

- 1) 根据关键词组合策略进行文本匹配分析，支持关键词与或非逻辑表达式
- 2) 抗空格、星号、下划线等特殊字符干扰
- 3) 单机支持>1000个组合策略

关键词权重

- 1) 对各关键词进行加权减权计算，根据文本的总分值进行总体判定
- 2) 滤除包含特定关键词的合法文本，有效降低误判率

处理性能

- 1) 文字处理性能70Mbps
- 2) 支持文本格式包括HTML、WML、TXT、DOC、RTF以及PDF

编码支持

- 1) ASCII, ISO-8859-1, GB2312, GBK, UTF-8, UTF-16
- 2) 中文、英文、维语、藏语

海量语料库

- 1) 积累超过1G的不良文本内容，用于自然语义分析算法训练

关键词情感度标注

- 1) 对不良文本常见关键词实现语义标注

中文智能分词

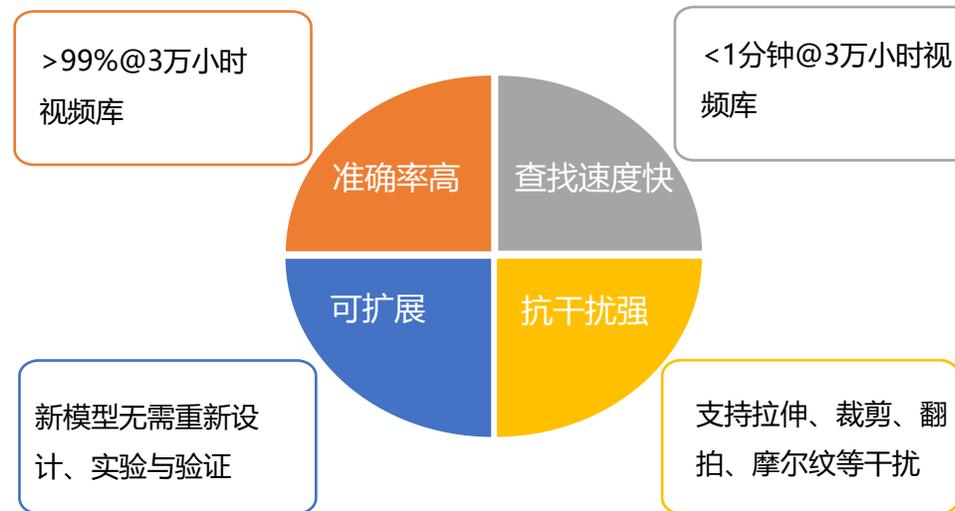
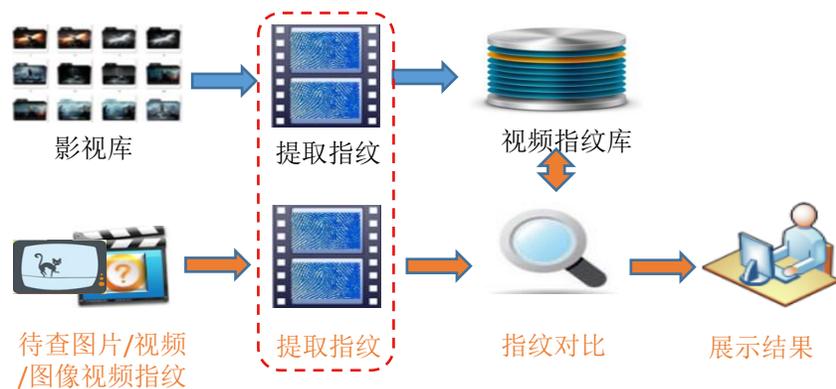
- 1) 自动编码转换；
- 2) 格式化文章内容，提取标题、正文、时间等信息
- 3) 去除特殊字符抗干扰

分类过滤

- 1) 根据配置规则进行语义分类
- 2) 区分色情类、恐怖类、暴力类、政治类、邪教类等分类文本

热词提取

- 1) 获取待处理的文章集合；使用朴素贝叶斯来获取到每篇文章的关键词及其权重；然后根据TF-IDF计算关键词的集合中出现的权重，根据权值排序，选出权值较大的N个词为热词，作为关键词策略备选



模式识别是指计算机通过算法自动识别图像或视频中的物品和标识等，通过计算机的识别训练过程会逐步提升识别的准确率。

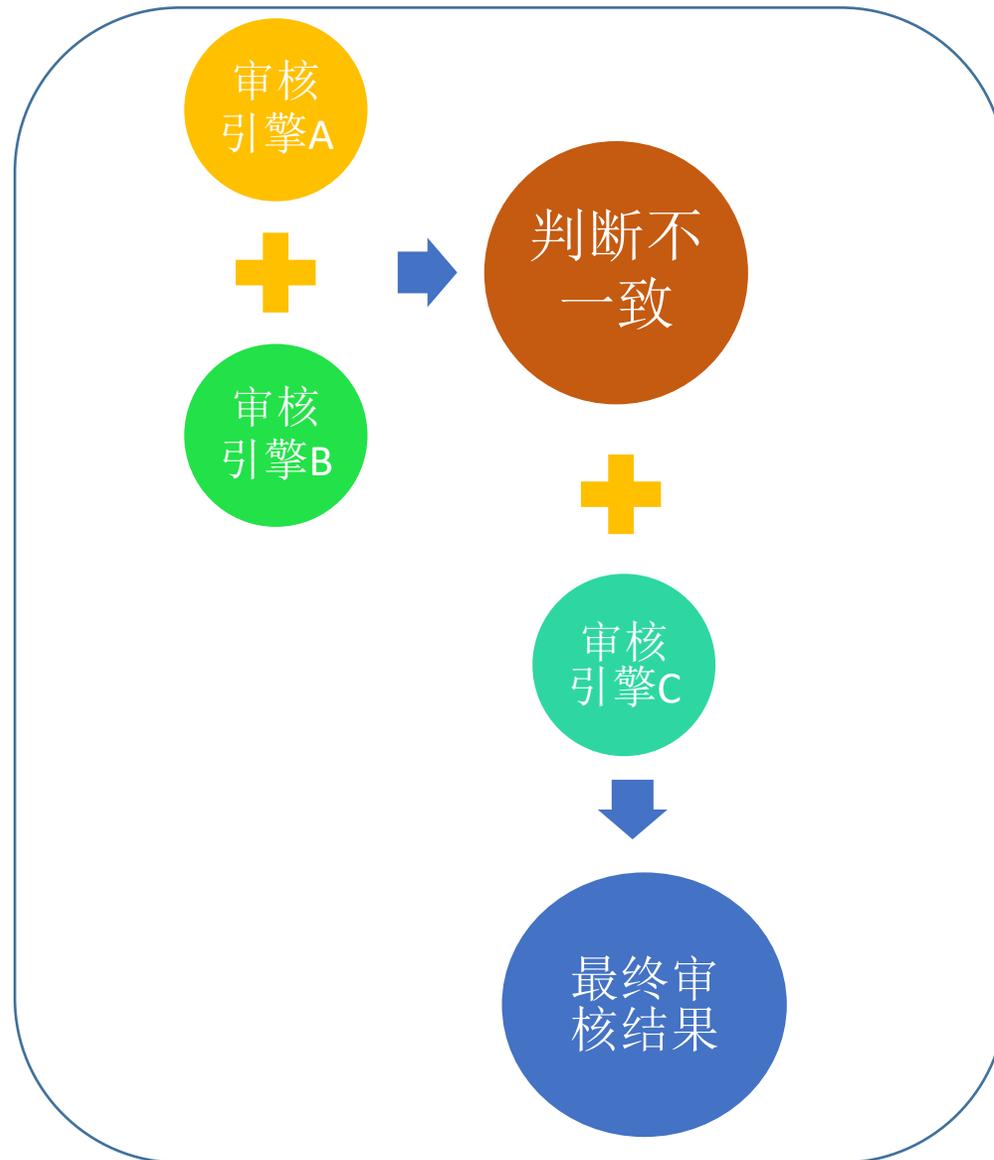
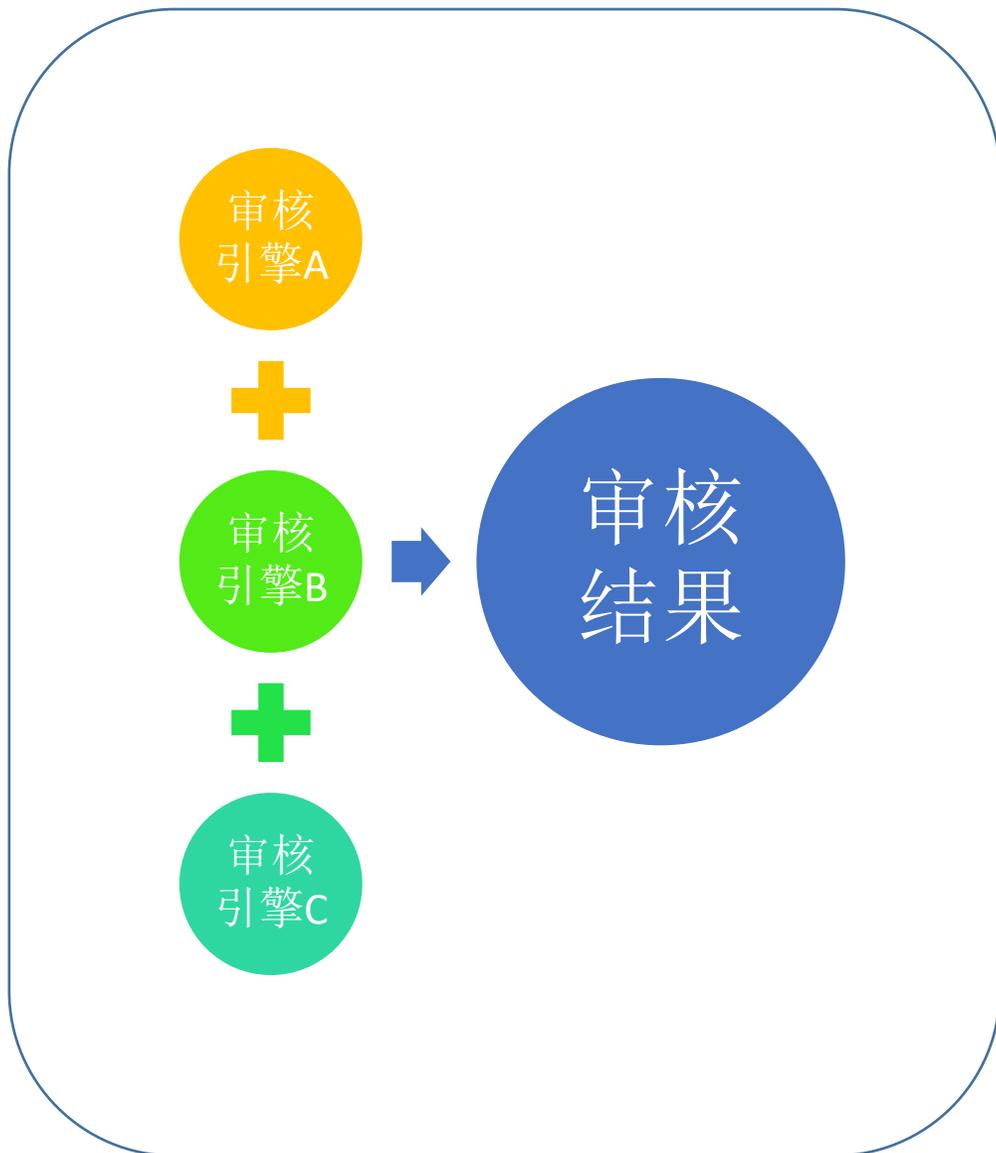
智能无参考识别

- 特定场景与物体识别



- 台标/LOGO识别

- 人物识别





审核员

考核

技能

举报人/志愿者

激励

评级

内容安全为人民，内容安全靠人民



■ 曝光范围区分

会把灰色内容限制在用户的个人主页或者同城曝光，这样就只有粉丝或者部分用户能看到，而且内容曝光量达到一定程度后就下掉，事后看没有半点痕迹。

■ 区分曝光人群

如新用户看不见，资深用户才曝光；女性不推送，男性才推送。

■ 区分曝光区域

如在一线城市和监管严格的城市不曝光而在其它区域曝光灰色内容。

■ 区分曝光时段

在深夜就容易搜索到灰色内容，而白天则更干净。

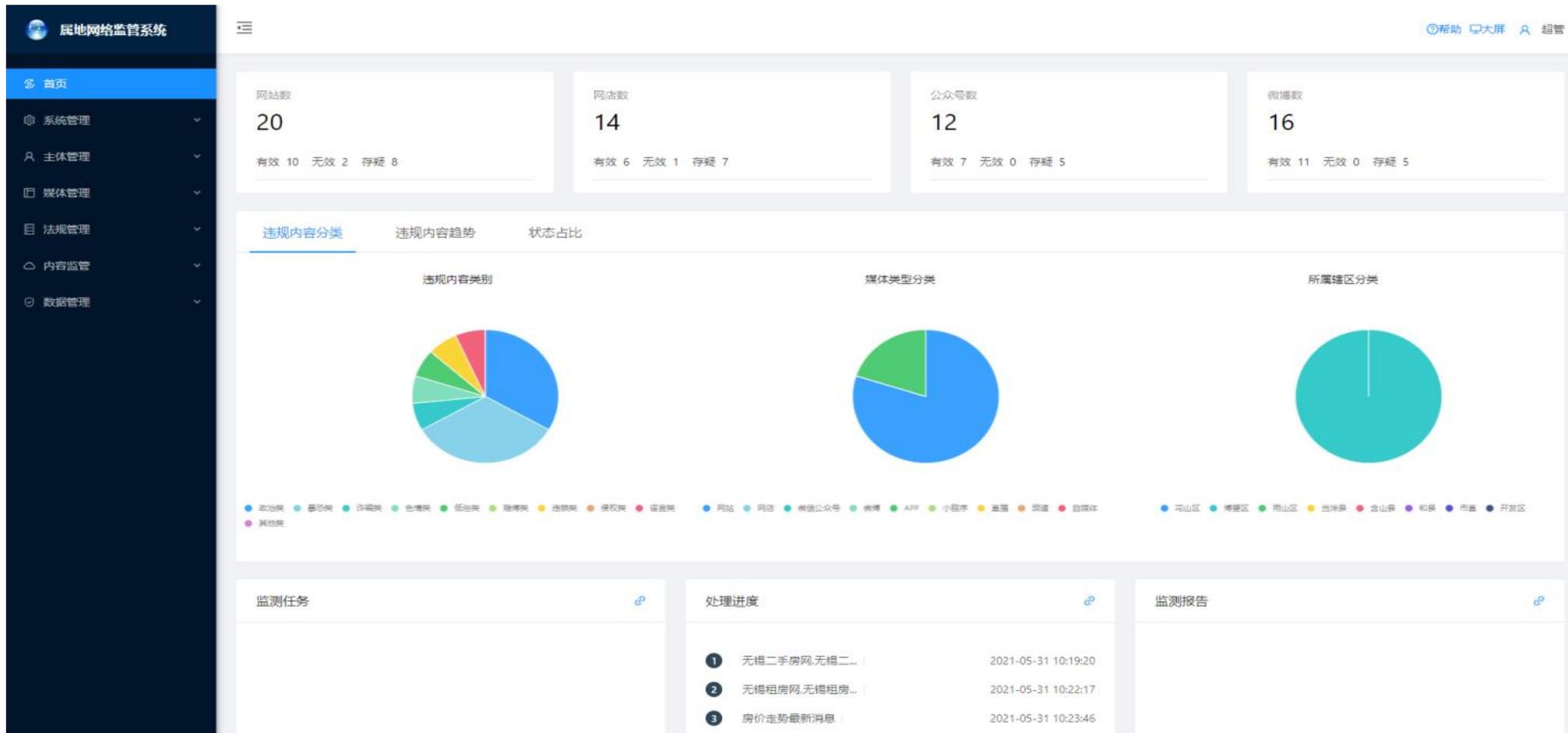
目录

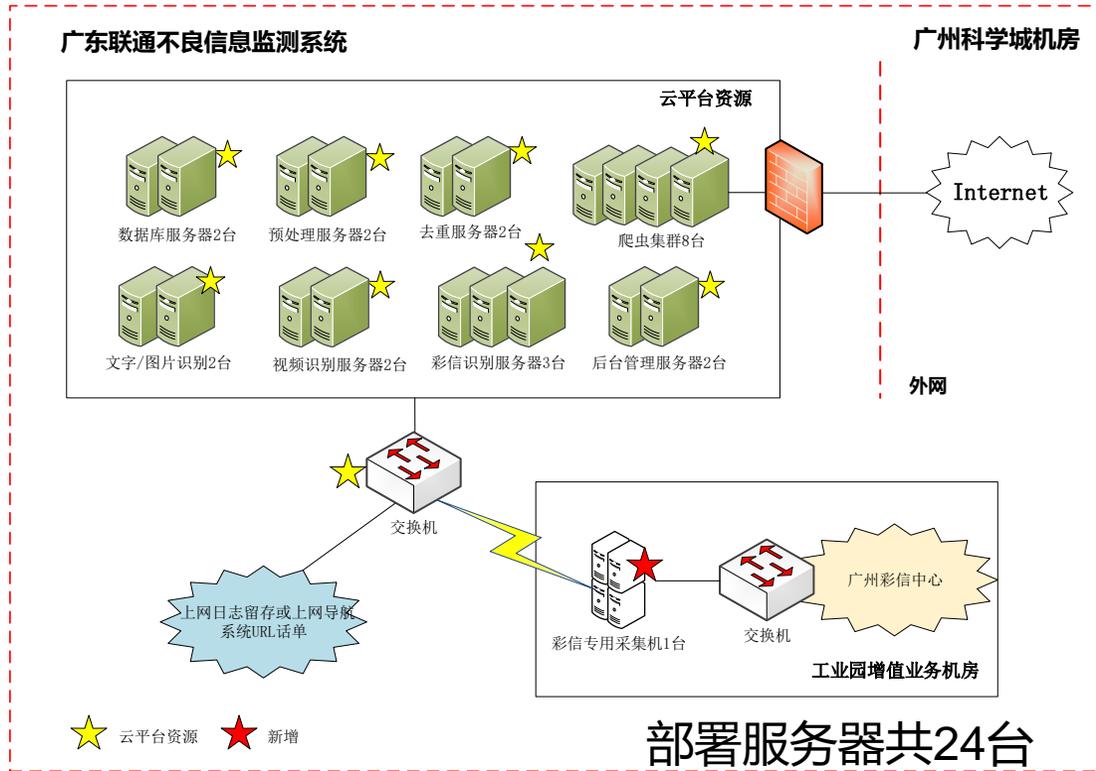
背景介绍

解决方案

案例分享







平均每天处理

- 爬取400万URL
- 识别文本30MB/秒
- 识别彩信200条/秒
- 识别图片80万张/天
- 视频40GB/天

- **监控目标:** XX联通所有WAP上网日志 16亿条/天, 去重后约400万条。
- **识别周期:** 每天进行实时数据的采集与分析。
- **数据获取方式:** 通过把采集到的WAPS上网日志中的URL通过自动接口和界面导入两种方式导入到系统中, 然后智能分发到爬虫模块进行采集
- **爬虫模块:** 负责对系统导入的URL数据进行通过爬虫技术进行数据采集。对于文本数据直接进行关键词过滤; 对于图像视频数据输出到图像视频识别模块进行识别
- **文字识别模块:** 系统通过高性能的关键词匹配算法对文本内容进行识别, 支持关键词权重设置, 支持关键词逻辑表达式、支持减权关键词功能。支持当前网络中所有主流的文本格式包括HTML、WML、TXT、DOC、RTF以及PDF
- **图像、视频识别模块:** 负责对爬虫的图像、视频数据进行违规识别, 提供基于指纹库的比对识别和基于智能图像分析的模式识别两种识别方式。支持对黄色、暴恐等不良内容的智能识别。

视频点播监控系统

- 个人中心
- 系统管理
- 系统监控
- 酒店信息
 - 酒店信息管理
 - 影片主页
 - 点播记录
 - 违规记录
- 供应商管理
- 报告管理

首页 / 酒店信息 / 影片主页

个人中心 酒店信息管理 影片主页

输入影片名称搜索 | 选择影片状态搜索 | 选择违规检测搜索 | 选择酒店名称搜索 | 选择供应商搜索 | 入库开始日期: 入库结束日期 | 搜索 | 重置



监护风云

主演: 蕾雅·德吕盖
入库时间: 2021-10-12 22:21:38
所属酒店: 南京金鹰国际酒店



叶问4: 完结篇

主演: 甄子丹/吴樾
入库时间: 2021-10-13 12:12:00
所属酒店: 南京金鹰国际酒店



哪吒之魔童降世

主演: 其他
入库时间: 2021-10-13 10:36:34
所属酒店: 南京金鹰国际酒店



好命先生

主演: 陈小春 / 安以轩 / 刘桦
入库时间: 2021-10-13 16:47:11
所属酒店: 南京金鹰国际酒店



我的青春期

主演: 包贝尔 / 孙怡 / 冯四
入库时间: 2021-10-13 16:08:21
所属酒店: 南京金鹰国际酒店



八佰

主演: 王千源 / 张译 / 姜武 / 黄...
入库时间: 2021-10-13 12:25:00
所属酒店: 南京金鹰国际酒店



网站数

3679

有效 2355 无效 1324 存疑 0

网店数

1016

有效 215 无效 287 存疑 514

公众号数

1010

有效 822 无效 188 存疑 0

微博数

854

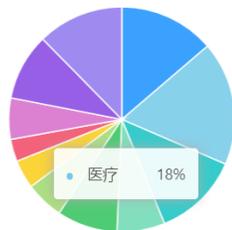
有效 854 无效 0 存疑 0

违规广告分类

违规广告趋势

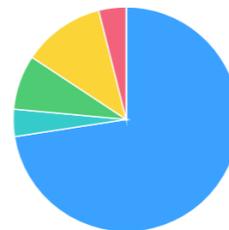
状态占比

违规广告类别



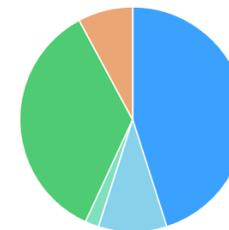
● 药品 ● 医疗 ● 医疗器械 ● 食品 ● 特殊医学用途配方食品 ● 农药
● 兽药 ● 烟草 ● 酒类 ● 化妆品 ● 房地产 ● 广告综合性法律、法规、规章
● 长江禁渔

媒体类型分类



● 网站 ● 网店 ● 微信公众号 ● 微博 ● APP ● 小程序 ● 直播

所属辖区分类



● 雨花分局 ● 赛虹桥分局 ● 西善桥分局 ● 板桥分局 ● 铁心桥分局
● 梅山分局 ● 古雄分局 ● 开发区分局 ● 非本辖区

监测任务



- | | | |
|---|---------|---------------------|
| 1 | 11 | 2021-10-23 03:25:13 |
| 2 | 10月专项监测 | 2021-10-11 11:16:53 |

处理进度



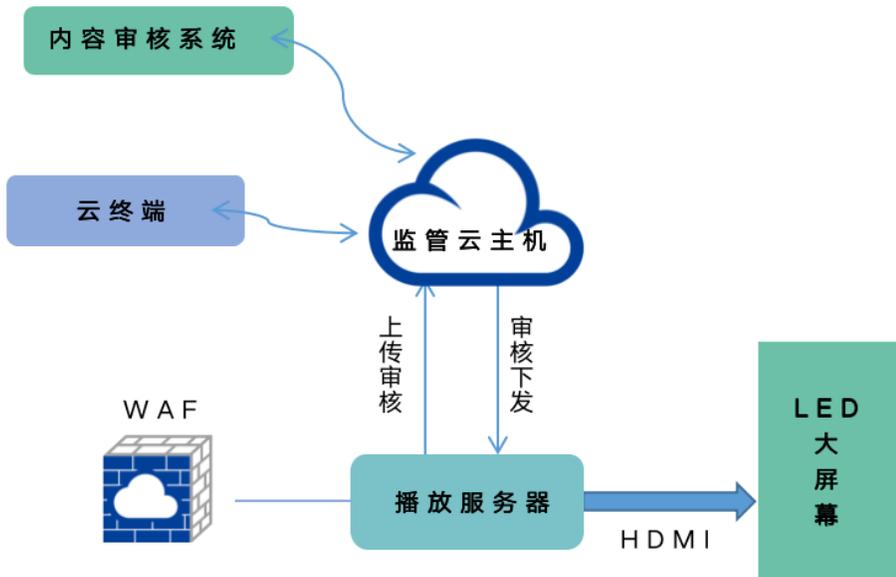
- | | | |
|---|-----------------------|---------------------|
| 1 | 硬不起来随便去药店买... 雨花分局 | 2021-03-29 09:54:53 |
| 2 | 南京哈宁轴承制造有限... 铁心桥分局 | |

监测报告



- | | | |
|---|--------------------|---------------------|
| 1 | 2022年8月广告监测报告.docx | 2022-08-31 23:59:02 |
| 2 | 2022年7月广告监测报告.docx | 2022-07-31 23:59:02 |

XX户外大屏监管系统



习近平出席党史学习教育动员大会并发表重要讲话

←

党史学习教育动员大会 20 日上午在北京召开。中共中央总书记、国家主席、中央军委主席习近平出席会议并发表重要讲话。他强调，在全党开展党史学习教育活动，是党中央立足党的百年历史新起点、统筹中华民族伟大复兴战略全局和世界百年未有之大变局、为动员全党全国满怀信心投身全面建设社会主义现代化国家而作出的重大决策。全党同志要做到学史明理、学史增信、学史崇德、学史力行，学党史、悟思想、开新局、办实事，以昂扬姿态奋力开启全面建设社会主义现代化国家新征程，以优异成绩纪念建党一百周年。←

中共中央政治局常委李克强、栗战书、汪洋、赵乐际、韩正，国家副主席王岐山出席会议，中共中央政治局常委王护宁主持会议。←

习近平强调，我们党历来重视党史学习教育，注重用党的奋斗历程和伟大成就鼓舞斗志、明确方向，用党的光荣传统和优良作风限定信念、凝聚力量，用党的实践创造和历史经验启迪智慧、砥砺品格。党的十八大以来，党中央高度重视学习党的历史，提出了一系列要求。在庆祝我们党百年华诞的重大时刻，在“两个百年”奋斗目标历史交汇的关键节点，在全党集中开展党史学习教育活动，正当其时，十分必要。←

习近平指出，我们党的一百年，是矢志践行初心使命的一百年，是筚路蓝缕奠基立业的一百年，是创造辉煌开辟未来的一百年。回望过往的奋斗路，眺望前方的奋进路，必须把党的历史学习好、总结好，把党的成功经验传承好、发扬好。在全党开展党史学习教育，是忘记初心使命、推进中华民族伟大复兴

网站CMS



智能审核
接口

- 检测时间: 2021-10-14

检测时间: 2021-10-14 11:16:26
错误词: “两个百年”奋斗目标
错误原因: 习惯表述
建议词: [“两个一百年”奋斗目标]
错误的严重程度: 一般错误
建议的纠错动作: 替换

答复 解决

- 检测时间: 2021-10-14

Thanks!

请各位领导和专家批评指正!