

**宁夏政务系统**

**数据抓取方案**

|  |  |  |  |
| --- | --- | --- | --- |
| 文档版本号： | v 1.0 | 文档编号： |  |
| 文档密级： | 保密 | 归属部门/项目： | 产品方案 |
| 产品名： | XX | 子系统名： |  |
| 编写人： | XX | 编写日期： | 2017-07-28 |

**深圳市茁壮网络股份有限公司 版权所有**

**内部资料注意保密**

**修订记录：**

|  |  |  |  |
| --- | --- | --- | --- |
| **文档版本** | **修订人** | **修订日期** | **修订描述** |
| V 1.0 | Zhangmin | 2017-07-28 | 初稿 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**目录**

[一、 需求说明 4](#_Toc32168)

[二、 素材来源 4](#_Toc32362)

[1、 政务网对接栏目 4](#_Toc5684)

[2、 自治区人大栏目对接 4](#_Toc18579)

[3、 自治区政协栏目对接 5](#_Toc11377)

[4、 智慧党建栏目对接 5](#_Toc5973)

[三、 需要抓取的内容 6](#_Toc15582)

[1、 文字类资源 6](#_Toc23703)

[2、 视频类资源 6](#_Toc17312)

[3、 领导介绍 6](#_Toc5721)

[4、 信件 7](#_Toc14177)

[5、 提案、建议 7](#_Toc13984)

[四、 接口说明 7](#_Toc18444)

# 需求说明

宁夏智慧党建项目中，党员学习素材需要从其他网站中爬取。

功能1：将爬取到的素材进行归类存储；

功能2：提供数据的接口供终端调用。

# 素材来源

## 政务网对接栏目

截取网址：<http://www.nx.gov.cn/>（宁夏回族自治区人民政府网）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **首页** | **自治区政府** | **政务公开** | **互动交流** | **塞上江南** | **专题数据** |
| 央网动态 | 自治区政府领导 | 自治区政府文件（图文，通用） | 主席信箱（main\_box） | 主要城市（ok） | 振奋精神  实干兴宁 |
| 政务动态（专题，直播，两会） | 执政实录（视频） | 信息公开指南 | 人大建议(info\_box表格) | 专题旅游（ok） | 权力和责任清单(固定文件下载) |
| 市县动态（ok） | 政府常务会议（ok） | 政府信息公开目录（） | 政协提案 | 精品线路（ok） | 民生实事（视频） |
| 政策解读(ok) | 政府工作报告（ok） | 政府信息公开年报（去掉） | 网络征集 |  | 脱贫攻坚（视频） |
| 通知公告(ok) |  | 政府信息公开制度 |  |  | 一带一路（视频） |
| 人事任免（ok） |  |  |  |  |  |

## 自治区人大栏目对接

截取网址：<http://www.nxrd.gov.cn/index.html>（[宁夏回族自治区人民代表大会常务委员会](http://www.so.com/link?m=aU5GjZukskJk2r8o0YTVk7Ir%2BjvqALlpgJCEuP%2BLMsvAXtZ92fC1ZG1d7s114KRn3JsVwkU95EfTO8nWa4JYTN%2FZolAhKFb6Kp3TMPyZIwjTIBAhLz9%2Fkcgnmmqs%3D" \t "_blank) ）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **人大概况** | **人大新闻** | **常委会工作** | **工作动态** | **人大专题** |
| 组织机构 |  | 工作计划(ok) | 通知公告（ok） | 全国人代会宁夏代表团(ok) |
| 领导视窗(简历图片问题) |  | 立法动态(ok) | 自身建设(ok) | 自治区人民代表大会会议(ok) |
| 领导讲话（TRS\_Editor） |  | 监督视窗（ok） | 代表工作(ok) | 自治区人大常委会会议 |
|  |  | 重大事项决定（ok） | 市县区人大(ok) | 自治区人大常委会主任会议(ok) |
|  |  | 人事任免（ok） |  | 机关主题教育实践活动 |

## 自治区政协栏目对接

截取网址：<http://www.nxzx.gov.cn/>（宁夏政协网）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **政协概况** | **政协会议（2017）** | **委员会工作** | **政协工作动态** | **政协新闻** | **政策法规** |
| 组织机构 | 政协全体会议(ok) | 提案委员会(ok) | 工作动态 | 热点专题（ok） | 规章制度 |
| 政协简介（ok） | 政协常委会议（ok） | 经济委员会(ok) | 市县政协 | 各地政协新闻（ok） | 理论研究 |
| 领导讲话（ok） | 政协主席会议（ok） | 人口资源环境委员会(ok) | 委员风采 | 图片动态（ok） | 文史资料 |
| 党派团体（ok） |  | 科教文卫体委员会(ok) | 社情民意 | 书画摄影 |  |
|  |  | 社会和法制委员会(ok) |  |  |  |
|  |  | 民族和宗教委员会(ok) |  |  |  |
|  |  | 文史和学习委员会(ok) |  |  |  |
|  |  | 港澳台侨联络委员会(ok) |  |  |  |

## 智慧党建栏目对接

截取网址：http://www.nxdjw.gov.cn/（宁夏党建网）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **党建动态** | **公示公告** | **远程教育** | **要闻动态** | **党建工作** | **专题网页** |
| 基层党建 | 公示公告 | 党员教育 | 党建要闻 | 干部工作 | 12380举报网站-去掉 |
| 党建研究 | 任前公示 | 共产党员电视栏目 | 高层声音 | 人才工作 | 组工资讯 |
|  | 人事任免 | 资源中心 | 党建新闻 | 资料下载 | 党史教育网 |
|  |  | 专题专栏 | 图片新闻 | 政策法规 | 反邪教警示教育 |
|  |  | 全国资源 |  | 廉政建设 | 宁夏干部教育培训网络学院 |
|  |  |  |  |  | 三农呼叫中心-去掉 |

# 需要抓取的内容

抓取到的素材根据上一章表格中的栏目结构进行分类存储；

## 文字类资源

文字类新闻需要抓取的信息包括：

文章标题（必选）

文章副标题（可选）

正文（必选）

图片链接（可选）

图片说明（可选）

视频链接（可选）

视频说明（可选）

新闻来源（可选）

新闻发表时间（可选）

第几届委员会（可选）

所属专题名称（可选）

## 视频类资源

视频类资源需要抓取的信息包括：

视频标题（必选）

视频链接（必选）

视频发表时间（必选）

## 领导介绍

领导介绍需要抓取的信息有：

领导姓名（必选）

领导职务（可选）

领导头像（必选）

工作分工（可选）

个人简介（必选）

## 信件

信件类新闻需要抓取的数据有：

来信人（必选）

来信时间（必选）

诉求类型（必选）

所属区域（必选）

来信标题（必选）

来信内容（必选）

回复部门（必选）

回复内容（必选）

回复时间（必选）

## 提案、建议

议案、建议类新闻需要抓取的数据有：

编号（必选）

届次（必选）

类别（必选）

主办（必选）

协办单位（必选）

案由（必选）

正文（必选）

办理情况（必选）

# 接口说明

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 功能 | 接口说明 | 备注 |
|  | 获取素材 | 根据栏目结构、媒资类型返回素材 |  |