[爬虫-keep-设计文档](http://doc.pajk-ent.com/pages/viewpage.action?pageId=67194496)

[转至元数据结尾](http://doc.pajk-ent.com/pages/viewpage.action?pageId=67194496#page-metadata-end)

* Created and last modified by [朱章春](http://doc.pajk-ent.com/display/~ZHUZHANGCHUN967) [大约6小时以前](http://doc.pajk-ent.com/pages/diffpagesbyversion.action?pageId=67194496&selectedPageVersions=18&selectedPageVersions=19)

[转至元数据起始](http://doc.pajk-ent.com/pages/viewpage.action?pageId=67194496#page-metadata-start)

爬虫对接

| **团队** | **对接人** | **联调** |
| --- | --- | --- |
| 爬虫 | 宗晓龙 |  |
| 内容 | 李新 | 4.27-4.29 |

接口地址：

[http://srv.test.pajk.cn/hl-sync/infox/receive?jsonParam=](http://srv.test.pajk.cn/hl-sync/haina/receive?jsonParam=){}

[http://](http://srv.test.pajk.cn/hl-sync/haina/receive?jsonParam=)[srv.jk.cn/hl-sync](http://srv.jk.cn/hl-sync)[/hl-sync/infox/receive?jsonParam=](http://srv.test.pajk.cn/hl-sync/haina/receive?jsonParam=)[{}](http://srv.jk.cn/hl-sync)

参数：

json对象如下：

private String pageUrl; //文章采集地址  
private String pageUrlDomain; //文章采集地址域名  
private String pageTitle; //网页标题  
private String articleTitle; //文章标题  
private String articleAuthor; //文章作者  
private String articleFrom; //文章来源网站  
private String articleContentFrom; //文章中写明的来源名称  
private String articlePublishDate; //文章发布时间  
private String articleContent; //文章正文  
private String articleTextContent; //文章纯文字正文  
private String articleAbstract; //文章摘要  
private String articleKeyword; //文章关键词  
private String articleFingerprint; //文章的语义指纹(唯一ID)  
private String tag; //文章的TAG词  
private String articleImageUrl; //文章中的图片URL  
private String article23Class; //文章分类(新华社23分类)  
private String watchSourceUrl;//资讯监控地址  
private String showType; //网易有料文章类型：article, video, svideo  
private int type; //文章类型，与头条类型保持一致  1:资讯文章  5:资讯视频  10:图文帖  11:小视频

接口开发：

1. 支持传健康号；
2. 支持小视频，帖子；
3. 标签 -(官方必读、用户经验)；
4. 导航-theme主题分类；class23
5. 是否为流媒体；m3u8;
6. 分类区分，外部ID对应；
7. 根据三级分类名，导航维护外部ID，用于课程跳转，关联帖子与导航关系；

设计：

本次需求入参对象新增如下字段：

private String outId;  //外部ID

private String outType; //外部类型

private long userId; //健康号

private String label; //如官方必读

private String feature; //扩展

private String ext1; //扩展1

private String ext2; //扩展2

private String ext3; //扩展3

流程：

1. 爬虫发布一次；
2. 调用rest接口测试+线上‘；

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 必填 |  |  |  | 必填，如 官方必读 |  | 同title |  | 必填 | ? keep | ? keep | 必填 | 带html标签富文本，必填 type =1 | ？摘要 | 摘要 | 有就传，可选 | 空 | 有就传，可选 | 必填，http原始地址 |  | 传三级导航名称，最好有名称 | 空 | 空 | 空 | 空 | 必填 ?前部分 |  | 1：富文本资讯 5：视频  10：图文帖  11：小视频 |  |
| 1 | <http://kandian.youth.cn/index/detail?sign=6d85Bb8Gyk4xAgV> |  |  |  |  |  | 儿子每月只寄钱不回家，母亲到处说他不孝，六岁孙子说爸爸早死了 | 儿子每月只寄钱不回家，母亲到处说他不孝，六岁孙子说爸爸早死了 |  | 中国青年网 | 中国青年网 | 2018-04-13 10:32:38.0 | <p>　　孙大妈是个农村的妇女，她一辈子没有出过村里的大门，但他有个很优秀的儿子，这也是孙大妈一辈子最骄傲的事。</p> <p>　　孙大妈的老伴在儿子十几岁的时候，生了一场重病就去世了，留下这孤苦伶仃的母子二人相依为命。孙大妈牢记老伴临走前的嘱托，一定要把儿子抚养成人，成为家族的骄傲。</p> <p>　　她虽然大字不识一个，但是她会在儿子学习的时候陪在一旁，夏天的时候，家里没有风扇，就拿着扇子给儿子吹风，冬天家里没有暖气，孙大妈就给儿子烧炕，让儿子趴在被窝里写作业，好在孙大妈儿子争气，从小学习成绩就很优秀。</p> <p>　　孙大妈是盼星星盼月亮，终于盼到儿子考上了名牌大学，村里人都投来羡慕的眼光，孙大妈恨不得带着儿子天天在大街里转悠，恨不得让街坊邻居跟在屁股后面夸赞他们母子二人。</p> <p>　　她白天风光，晚上也有落泪的时候，儿子要上大学，那就得需要巨额的学费和生活费，孙大妈一个妇道人家，怎么能拿得出那么多钱。儿子似乎看出了妈妈的担忧，在升大学的暑假，去了城里打工赚钱。孙大妈也在附近工厂加班加点干活，终于在儿子上大学前把钱给凑齐了。</p> <p>　　她儿子体谅家里困难，在大学拼命学习，把学校的各项奖学金统统拿了下来，孙大妈也是看到了生活的希望。大学四年结束后，儿子顺利被一家大公司给录用了，这又成了孙大妈在街坊邻居面前吹捧的事情了。</p> <p><!--分页 --></p> <p>　　儿子把每个月的工资都会拿出一部分寄给孙大妈，孙大妈不用再每天卖命的干活了。村里也有好多媒婆争着抢着给孙大妈儿子说媳妇，孙大妈想尊重儿子意愿，就打电话询问儿子情感状况，但没想到儿子竟然说有女朋友，而且要带回家给她看看。孙大妈自然是高兴不得了。</p> <p>　　儿子带着女友回来，这姑娘是城里人，很懂礼貌，孙大妈也是越看越喜欢，就同意两个人的婚事了。儿子结婚之后，孙大妈就开始催儿子赶快生孩子，她要趁自己还能动，给儿子照料孩子。</p> <p>　　孙大妈的儿子遵照母亲的建议，在结婚第二年就生了孩子，还是个大胖小子，这小可是把孙大妈给乐坏了。在街里逢人就说自己当奶奶了，有孙子可以抱了。村子人都很羡慕嫉妒她，儿子有出息，在大城市上班。还娶了一个城里姑娘，现在又有了孙子，人生赢家啊！</p> <p>　　就这样过了几年，每到寒暑假孙大妈都会把孙子接回来玩一段时间，过年的时候孙大妈也会去城里住一段时间，一家四口生活的很幸福。在孙子五岁那年，家里出了点状况，孙大妈发现儿子已经很久没给家里打电话了，每个月只是按时寄钱，就连五一，十月一这种小长假儿子也没有回来转一转，孙大妈打电话给儿子，每次都是儿媳接的电话，说儿子在忙不方便接，更让孙大妈疑惑的是，自己孙子寒暑假也没有回来，孙大妈开始不满儿子的做法，就开始跟街坊邻居诉苦，说儿子变了，有钱了就忘了娘，不孝顺了，村里人心里也暗暗嘲讽，说孙大妈当初夸儿子夸的那么好，现在又开始唱反调，这就是风水轮流转啊！</p> <p><!--分页 --></p> <p>　　眼看孙子六岁生日到了，孙大妈决定去城里一趟，孙大妈先去孙子学校等孙子放学，接到孙子的时候孙大妈就迫不及待问儿子的事情，但没想到孙子竟然淡淡地说：“我爸一年前就死了啊，出差的时候发生了车祸！”</p> <p>　　孙大妈听完这番话，整个人都傻了，儿子竟然早就死了，那每个月的钱怎么还会按时寄给自己？孙大妈才想明白，都是儿媳妇在瞒着自己，是怕自己接受不了这个事情，孙大妈越想越难过，当初还埋怨儿子不孝顺，现在人已经不在了。</p> <p>　　孙大妈抱着孙子就嚎啕大哭，她不知道自己该怎么面对这个事情！</p> | 孙大妈是个农村的妇女，她一辈子没有出过村里的大门，但他有个很优秀的儿子，这也是孙大妈一辈子最骄傲的事。 　　孙大妈的老伴在儿子十几岁的时候，生了一场重病就去世了，留下这孤苦伶仃的母子二人相依为命。孙大妈牢记老伴临走前的嘱托，一定要把儿子抚养成人，成为家族的骄傲。 　　她虽然大字不识一个，但是她会在儿子学习的时候陪在一旁，夏天的时候，家里没有风扇，就拿着扇子给儿子吹风，冬天家里没有暖气，孙大妈就给儿子烧炕，让儿子趴在被窝里写作业，好在孙大妈儿子争气，从小学习成绩就很优秀。 　　孙大妈是盼星星盼月亮，终于盼到儿子考上了名牌大学，村里人都投来羡慕的眼光，孙大妈恨不得带着儿子天天在大街里转悠，恨不得让街坊邻居跟在屁股后面夸赞他们母子二人。 　　她白天风光，晚上也有落泪的时候，儿子要上大学，那就得需要巨额的学费和生活费，孙大妈一个妇道人家，怎么能拿得出那么多钱。儿子似乎看出了妈妈的担忧，在升大学的暑假，去了城里打工赚钱。孙大妈也在附近工厂加班加点干活，终于在儿子上大学前把钱给凑齐了。 　　她儿子体谅家里困难，在大学拼命学习，把学校的各项奖学金统统拿了下来，孙大妈也是看到了生活的希望。大学四年结束后，儿子顺利被一家大公司给录用了，这又成了孙大妈在街坊邻居面前吹捧的事情了。 　　儿子把每个月的工资都会拿出一部分寄给孙大妈，孙大妈不用再每天卖命的干活了。村里也有好多媒婆争着抢着给孙大妈儿子说媳妇，孙大妈想尊重儿子意愿，就打电话询问儿子情感状况，但没想到儿子竟然说有女朋友，而且要带回家给她看看。孙大妈自然是高兴不得了。 　　儿子带着女友回来，这姑娘是城里人，很懂礼貌，孙大妈也是越看越喜欢，就同意两个人的婚事了。儿子结婚之后，孙大妈就开始催儿子赶快生孩子，她要趁自己还能动，给儿子照料孩子。 　　孙大妈的儿子遵照母亲的建议，在结婚第二年就生了孩子，还是个大胖小子，这小可是把孙大妈给乐坏了。在街里逢人就说自己当奶奶了，有孙子可以抱了。村子人都很羡慕嫉妒她，儿子有出息，在大城市上班。还娶了一个城里姑娘，现在又有了孙子，人生赢家啊！ 　　就这样过了几年，每到寒暑假孙大妈都会把孙子接回来玩一段时间，过年的时候孙大妈也会去城里住一段时间，一家四口生活的很幸福。在孙子五岁那年，家里出了点状况，孙大妈发现儿子已经很久没给家里打电话了，每个月只是按时寄钱，就连五一，十月一这种小长假儿子也没有回来转一转，孙大妈打电话给儿子，每次都是儿媳接的电话，说儿子在忙不方便接，更让孙大妈疑惑的是，自己孙子寒暑假也没有回来，孙大妈开始不满儿子的做法，就开始跟街坊邻居诉苦，说儿子变了，有钱了就忘了娘，不孝顺了，村里人心里也暗暗嘲讽，说孙大妈当初夸儿子夸的那么好，现在又开始唱反调，这就是风水轮流转啊！ 　　眼看孙子六岁生日到了，孙大妈决定去城里一趟，孙大妈先去孙子学校等孙子放学，接到孙子的时候孙大妈就迫不及待问儿子的事情，但没想到孙子竟然淡淡地说：“我爸一年前就死了啊，出差的时候发生了车祸！” 　　孙大妈听完这番话，整个人都傻了，儿子竟然早就死了，那每个月的钱怎么还会按时寄给自己？孙大妈才想明白，都是儿媳妇在瞒着自己，是怕自己接受不了这个事情，孙大妈越想越难过，当初还埋怨儿子不孝顺，现在人已经不在了。 　　孙大妈抱着孙子就嚎啕大哭，她不知道自己该怎么面对这个事情！ | 孙大妈是个农村的妇女，她一辈子没有出过村里的大门，但他有个很优秀的儿子，这也是孙大妈一辈子最骄傲的事。她儿子体谅家里困难，在大学拼命学习，把学校的各项奖学金统统拿了下来，孙大妈也是看到了生活的希望。 | 儿子,寄钱,早死,母亲,孙子,名牌大学,大学四年,妇道,有女朋友,奖学金 | 18045610745851760977 | 孙子/NR/0.304769 孙/NT/0.184004 扇子/NT/0.068167 风扇/NT/0.066768 星星/NT/0.051221 幸福/NT/0.034451 |  | 社会/财政，金融 |  | 10 |  | 2018-04-13 10:32:39.0 | 2018-04-13 10:32:39.0 | <http://demo.hylanda.com/download/youth_kd_mw.php> |  | 1 |  |
| 2 | <http://kandian.youth.cn/index/detail?sign=voX0mJ0raVDZQgy> |  |  |  |  |  | 天已微凉，守候相望，用45°角仰望你的伪装！将曾经的情殇，永世埋葬！ | 天已微凉，守候相望，用45°角仰望你的伪装！将曾经的情殇，永世埋葬！ |  | 中国青年网 | 中国青年网 | 2018-04-13 10:47:34.0 | <p>　　我，就像一个可有可无的影子，和寂寞交换着彼此的心事。</p> <p>　　对爱我无计可施，这无味的日子，眼泪，是唯一的奢侈。</p> <p>　　如何能见梦中事，哪怕一首打油诗。</p> <p></p> <p>　　柳柔千丝戏春燕，扬花如雪愁心间！风起声雀哀跃舞，青丝泛白垂眸前。</p> <p>　　莺雀过海寻比翼，匆匆一世难寻觅！曾经豆蔻天地老，奈何垂暮如约到。</p> <p>　　西风凉，情断肠，蓬莱仙阁在何方？欲将惆怅化心间，又恐此物污心田。<br></p> <p></p> <p>　　起初不相识，最终不相认。等待着你的关心，却直到关上了心。</p> <p>　　你心中的四季早已变幻，而我还守候在相识的季节。<br></p> <p>　　想你的时候会有幸福，但这幸福已不再拥有。</p> <p></p> <p>　　当泪流了下来，才明白，分离其实是另一种爱的表达。<br></p> <p>　　雨淋湿的空气，冻结了我的心，将美丽的童话，改成最妙的笑话。<br></p> <p>　　明明已经不再爱你，闭上眼，以为能够忘记，但流下的眼泪，却没有骗到自己……</p> <p>　　旧伤不及盖，新伤却再来。这样的伤痛，不想再来。<br></p> <p>　　期盼时间去淡化，却无奈，它已化为彼岸花。<br></p> | 我，就像一个可有可无的影子，和寂寞交换着彼此的心事。 　　对爱我无计可施，这无味的日子，眼泪，是唯一的奢侈。 　　如何能见梦中事，哪怕一首打油诗。 　　柳柔千丝戏春燕，扬花如雪愁心间！风起声雀哀跃舞，青丝泛白垂眸前。 　　莺雀过海寻比翼，匆匆一世难寻觅！曾经豆蔻天地老，奈何垂暮如约到。 　　西风凉，情断肠，蓬莱仙阁在何方？欲将惆怅化心间，又恐此物污心田。 　　起初不相识，最终不相认。等待着你的关心，却直到关上了心。 　　你心中的四季早已变幻，而我还守候在相识的季节。 　　想你的时候会有幸福，但这幸福已不再拥有。 　　当泪流了下来，才明白，分离其实是另一种爱的表达。 　　雨淋湿的空气，冻结了我的心，将美丽的童话，改成最妙的笑话。 　　明明已经不再爱你，闭上眼，以为能够忘记，但流下的眼泪，却没有骗到自己…… 　　旧伤不及盖，新伤却再来。这样的伤痛，不想再来。 　　期盼时间去淡化，却无奈，它已化为彼岸花。 | 对爱我无计可施，这无味的日子，眼泪，是唯一的奢侈。雨淋湿的空气，冻结了我的心，将美丽的童话，改成最妙的笑话。明明已经不再爱你，闭上眼，以为能够忘记，但流下的眼泪，却没有骗到自己……　　旧伤不及盖，新伤却再来。 | 眼泪,相识,美丽的童话,闭上眼,扬花,想你的时候,情殇,彼岸花,断肠,青丝 | 10895859252890905370 | 柳柔/NR/0.143370 豆蔻/NT/0.140441 幸福/NT/0.092085 |  | 文学，艺术 |  | 10 |  | 2018-04-13 10:47:32.0 | 2018-04-13 10:47:32.0 | <http://demo.hylanda.com/download/youth_kd_mw.php> |  | 11:小视频，轮播  5：长视频，下部分评论 |  |
| **id** | **page\_url** | **user\_id** | **out\_type** | **out\_id** | **label** | **page\_url\_domain** | **page\_title** | **article\_title** | **article\_author** | **article\_from** | **article\_content\_from** | **article\_publish\_date** | **article\_content** | **article\_text\_content** | **article\_abstract** | **article\_keyword** | **article\_fingerprint** | **tag** | **article\_image\_url** | **article\_23class** | **article\_class** | **status** | **reason** | **gmt\_created** | **gmt\_modified** | **watch\_source\_url** | **show\_type** | **type** | **feature** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 6217860 | <https://yiyouliao.com/api-server/rss/gooddoctor/item/recommendation/VL00MKJDVD3FAQZ.html> |  | 爸爸妈妈用新创意“惩罚”大头儿，结果发现不是什么创意都是对的 | 爸爸妈妈用新创意“惩罚”大头儿，结果发现不是什么创意都是对的 |  | [yiyouliao.com](http://yiyouliao.com/) | 菠萝联盟 | 2020-04-23 15:54:30.0 | <p style="TEXT-JUSTIFY: distribute; TEXT-ALIGN: justify; TEXT-INDENT: 30px; MARGIN: 0px 3px 15px" align="justify"></p> | <https://pingandoctor-video-nos.yiyouliao.com/icvt-20200423-4d228fae250e1796eaefb75436bf1622.mp4?time=1587630137&signature=2021DC82FFA6F3481C9AF305328716AD> |  | 创意;大头儿;爸爸妈妈;结果 |  | 大头儿 妈妈 创意 爸爸 | <https://pingandoctor-img-nos.yiyouliao.com/inforec-20200423-cd02690b75e9c91a88d8cf8bbd8a6c60.jpg?time=1587630137&signature=9DB1FC2D6A63377EEB57854472BD16E4> | 育儿 | 100 | 成功 | 2020-04-23 16:22:24.0 | 2020-04-23 16:22:24.0 | <https://yiyouliao.com/rss/common/gooddoctor/list/json?cat=beiyun> | video | 5 |  |
| 6217857 | <https://yiyouliao.com/api-server/rss/gooddoctor/item/recommendation/VL01IQ3NI18AVR9.html> |  | 女人对你释放三个信号，你就不要白费力气了，情商高男人秒懂！ | 女人对你释放三个信号，你就不要白费力气了，情商高男人秒懂！ |  | [yiyouliao.com](http://yiyouliao.com/) | 菠萝联盟 | 2020-04-23 15:54:56.0 | <p style="TEXT-JUSTIFY: distribute; TEXT-ALIGN: justify; TEXT-INDENT: 30px; MARGIN: 0px 3px 15px" align="justify"></p> | <https://pingandoctor-video-nos.yiyouliao.com/icvt-20200423-fa584367646f076ed7ede8b9ecc3eb9d.mp4?time=1587630137&signature=13B2FC41EDBB4F385E2E23271CB60AAB> |  | 女人;情商;男人;信号;力气;三个 |  | 商高 信号 女人 男人 | <https://pingandoctor-img-nos.yiyouliao.com/inforec-20200423-8addeae60b24393f1615216af1eea337.jpg?time=1587630137&signature=036DEBAD11127633481D663458A4BB85> | 情感-两性 | 10 |  | 2020-04-23 16:22:21.0 | 2020-04-23 16:22:21.0 | <https://yiyouliao.com/rss/common/gooddoctor/list/json?cat=liangxingqinggan> | video | 5 |  |
| **id** | **page\_url** | **page\_url\_domain** | **page\_title** | **article\_title** | **article\_author** | **article\_from** | **article\_content\_from** | **article\_publish\_date** | **article\_content** | **article\_text\_content** | **article\_abstract** | **article\_keyword** | **article\_fingerprint** | **tag** | **article\_image\_url** | **article\_23class** | **status** | **reason** | **gmt\_created** | **gmt\_modified** | **watch\_source\_url** | **show\_type** | **type** | **feature** |

[赞](http://doc.pajk-ent.com/pages/viewpage.action?pageId=67194496)成为第一个赞同者

* 无标签
* [编辑标签](http://doc.pajk-ent.com/pages/viewpage.action?pageId=67194496)

窗体顶端

编写评论...

窗体底端