20 廊坊市发改委

Title有问题

Id=1421485 content有前缀

22衡水市发改委

只采集了19年5月之前的记录，请检查配置是否有问题

39 忻州市政府

部分记录content为空，如id=1247341

51 赤峰市政府

publish\_org='发布时间'

56 鄂尔多斯市发改委

Abstract

76 锦州市发改委

Content不用采集“上一篇：XXX下一篇：XXX”

131 南京市政府

Content 不用采集“【打印本页】”等

徐州市政府

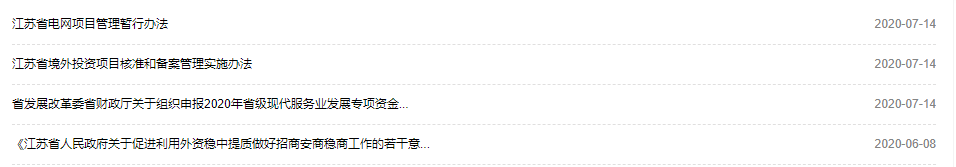
Content前缀

徐州市发改委

Publish\_org 不用加“来源：”

连云港发改委

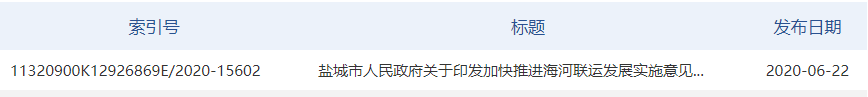
网站头几条记录未采集



盐城市政府

Content不用采集后缀“打印 关闭”

缺失第一条



临沂市发改委

如果可以的话，把content=‘如果您无法在线浏览此 PDF 文件，则可以

下载免费小巧的 福昕(Foxit) PDF 阅读器,安装后即可在线浏览  或

下载免费的 Adobe Reader PDF 阅读器,安装后即可在线浏览  或

下载此 PDF 文件‘删掉吧

深圳市发改委

很多content原网站没有附件，但是content的后缀带有“附件：”

珠海市政府

Content末尾不需要采集“分享到： 0“

江门市发改委

Content没有附件，后缀带有“附件下载：“

清远市发改委

Content 没有附件，后缀带有“相关稿件：”

清远市政府

Content没有附件，后缀带有“附件下载：”

潮州市发改委

Content没有附件，后缀带有“附件：”

潮州市政府

Content没有附件，后缀带有“附件：”

揭阳市发改委

Content没有附件或相关阅读时，后缀带有“相关附件： 相关阅读：”

哈密市政府

缺失网站前几条资讯



青岛市政府

若干记录采集了列表页的title，带省略号



应当采集详情页的title

青岛市发改委

Title日期后缀



烟台：网站504 待核验

菏泽市发改委

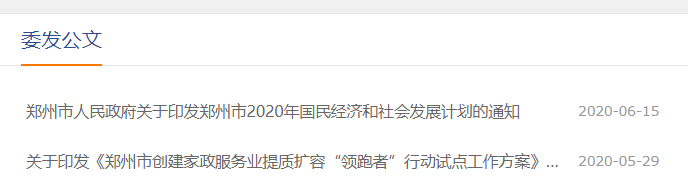
Title带换行



应当采集详情页的title

郑州市发改委

缺失第二条信息



惠州市政府

网页需从首页跳转；content前缀后缀问题；采集内容与配置内容不同

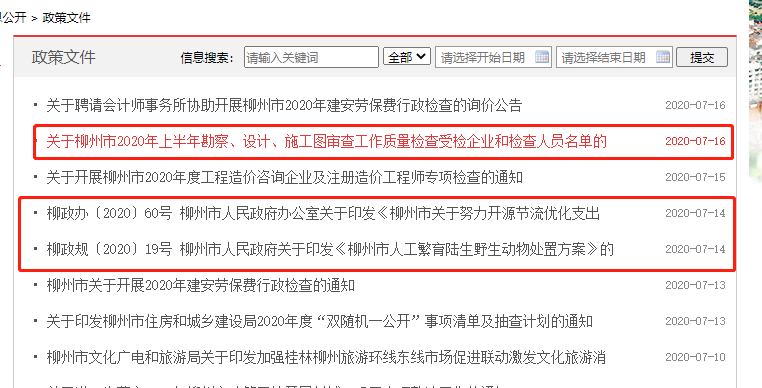




惠州市发改委：网页无法打开，待核验

柳州市政府

部分记录缺失



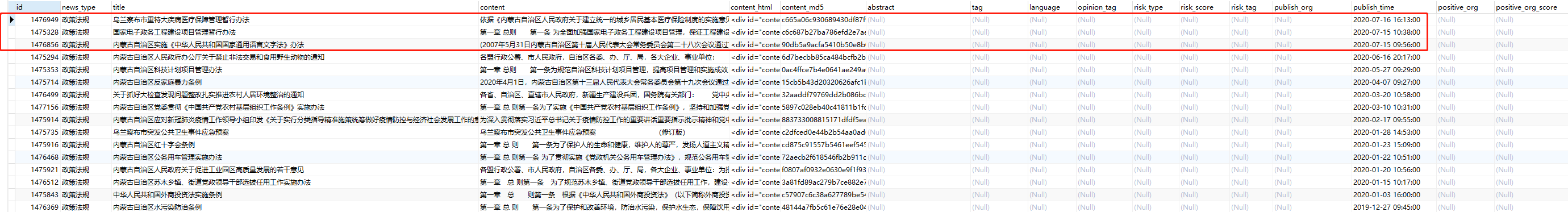
柳州市发改委

缺失2020年6月5号往后的内容，约前八页

乌兰察市政府

前三条记录是从哪儿采的呢





鞍山市发改委

这些记录对应的网页时间都是六月份或更早



梧州市政府

Content原网页无关联文件，但是后缀带“文件下载： 关联文件：”

梧州市发改委

Content原网页无关联文件，但是后缀带“文件下载： 关联文件：”

定西市政府

缺失部分记录，如



阜新市政府

Publish\_org有问题；content原网站均带有前缀“zzzz”，配置的时候去一下



铁岭市政府

Publish\_time和网页记录对应不上





衢州市发改委

缺失若干记录



舟山市政府

Content前缀后缀

芜湖：

Content后缀带“解读”

贵港市政府

Content后缀“文件下载：关联文件：”

玉林市政府

Content后缀“文件下载：关联文件：”

玉林市发改委

Content后缀“文件下载：关联文件：”

百色市政府

Content后缀“文件下载：关联文件：”

百色市发改委

Content后缀“文件下载：关联文件：”

山南市发改委

Content后缀“扫一扫在手机打开当前页面”

固原市政府

Content后缀“扫一扫在手机打开当前页面”

长春市政府

Content后缀

编 辑:

长春市人民政府 www.changchun.gov.cn

时间：2017-10-26 09:29

来源：

打印|关闭

常州市发改委

缺失7月1号往后的记录

（可以参考配置这个网页<http://www.changzhou.gov.cn/search/detail.php>，选择部门发改委）

南通市发改委

content后缀“打印本页 关闭窗口“

潍坊市政府

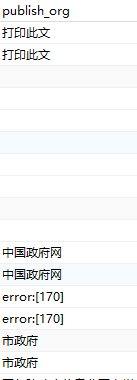
Title后缀“潍政发【xxxx】x号”

日照市发改委

News\_test表里为空

开封市政府

Publish\_org

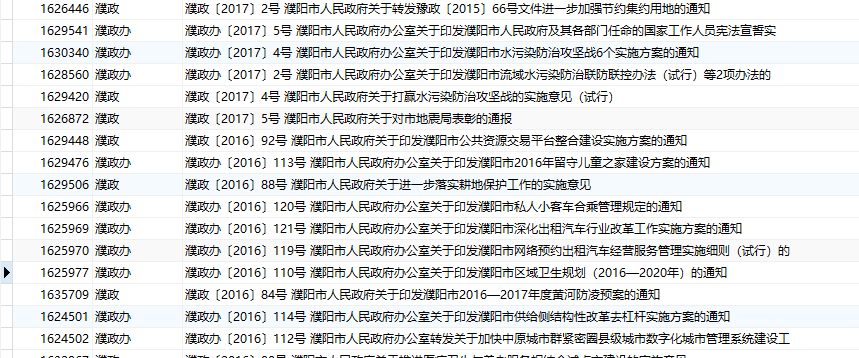


焦作市发改委

News\_test表里为空

濮阳市政府

部分title被截断，如id=1625970



贺州、河池、来宾

Content后缀 “文件下载： 关联文件：”