凤凰网网站分析（<http://www.ifeng.com/> ）

1. 专题类：<http://news.ifeng.com/mainland/special/xjpfwhskst/>

<http://d.news.ifeng.com/pc/special/40119/index.shtml>

<http://gongyi.ifeng.com/listpage/27451/1/list.shtml>

1. 新闻内容类：.shtml结尾网页 图解新闻（新闻正文内容为图片）是否抓取？

例：<http://news.ifeng.com/a/20170609/51216419_0.shtml>

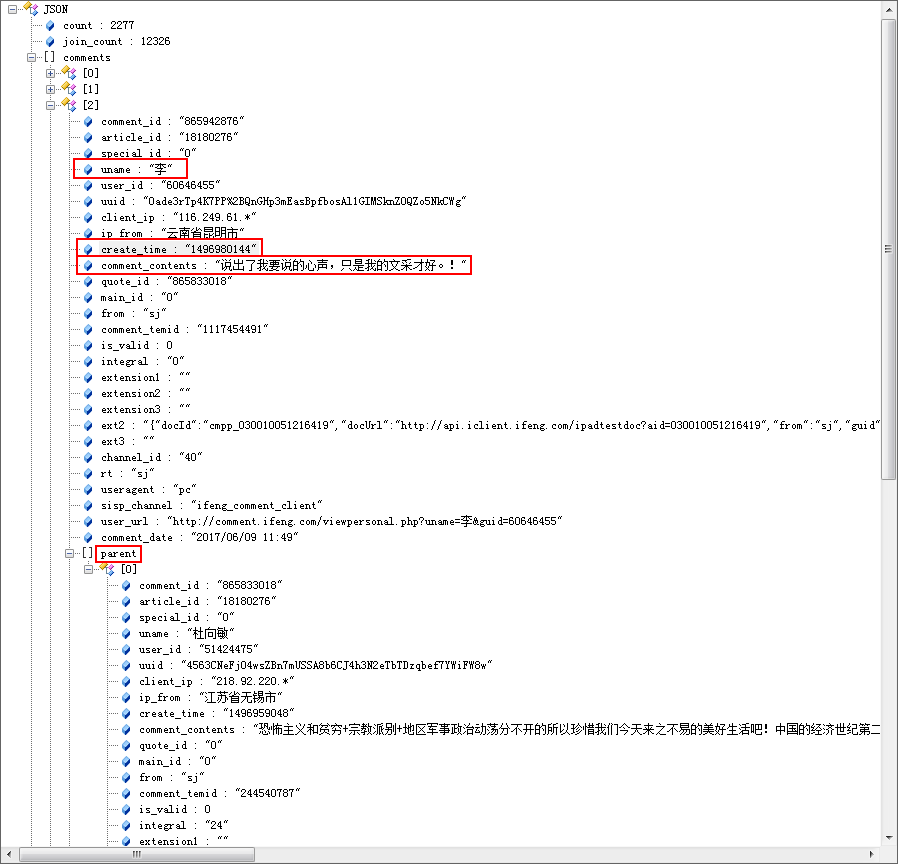
* 1. 标题：//\*[@id="artical"]/h1
  2. 发布时间：//\*[@class="p\_time"]/span[1] 日期格式：2017年06月09日 03:17:53
  3. 媒体名称：//\*[@class="p\_time"]/span/span
  4. 正文：//\*[@id="main\_content"]/p
  5. 评论获取方法：

每次最多获取20条评论内容，获取评论的url：

http://comment.ifeng.com/get.php?orderby=&docUrl={url}&format=js&job=1&p=1&pageSize=20，其中参数url为新闻的url。

返回的数据格式如下图所示：



将数据格式化，并对数据内容进行Unicode编码转换，取commentJsonVarStr\_\_\_中的jason数据，格式化及解码后的内容，如下图所示：

其中，评论数：count，评论内容：comments[comment\_contents]，评论时间毫秒数：comment[create\_time]，评论人昵称：comment[uname]，如果parent节点有值，则说明该条评论为parent评论的回复评论。

例：<http://news.ifeng.com/a/20170609/51217665_0.shtml>

1. 标题：//\*[@class="yc\_tit"]/h1
2. 发布时间：//\*[@class="yc\_tit"]/p/span[1] 日期格式：2017-06-09 08:09:17
3. 媒体名称：normalize-space(//\*[@class="yc\_tit"]/p/a)
4. 正文：//\*[@class="yc\_con\_txt"]/p
5. 评论获取方法：同上
6. 图片类：

例：<http://news.ifeng.com/a/20170609/51217493_0.shtml#p=1>

* 1. 标题：//\*[@class="titL"]/h1
  2. 发布时间：//\*[@class="titL"]/p 日期格式：2017.06.09 07:49
  3. 媒体名称：无
  4. 正文：//\*[@class="photoDesc"]
  5. 评论获取方法：同新闻类

例：<http://house.ifeng.com/pic/2017_06/09/38882585_0.shtml#p=1>

* 1. 标题：//\*[@id="title"]
  2. 发布时间：无
  3. 媒体名称：无
  4. 正文：//\*[@id="photo\_info"]/p
  5. 评论获取方法：同新闻类

1. 评论类：

例：<http://pl.ifeng.com/a/20170609/51217438_0.shtml>

* 1. 标题：//\*[@class="Atitle wrapIphone"]/h1
  2. 发布时间：//\*[@class="Slogo wrapIphone"]/p/span[2] 日期格式：2017.06.09
  3. 媒体名称：//\*[@class="Slogo wrapIphone"]/p/span/a
  4. 正文：//\*[@class="wrapIphone AtxtType01"]/p
  5. 评论获取方法：同新闻类

1. 凤凰论：

例：<http://pl.ifeng.com/opinion/fenghuanglun/276/1.shtml>

* 1. 标题：normalize-space(//\*[@class="artic\_tit"]/h1)
  2. 发布时间：substring(//\*[@class="artic\_tit"]/ul/li[1],0,11) 日期格式：2017.06.09
  3. 媒体名称：substring(//\*[@class="artic\_tit"]/ul/li[1],15)
  4. 正文：//\*[@class="artic\_list"]/p
  5. 评论获取方法：同新闻类

1. 高见：

例：<http://pl.ifeng.com/opinion/gaojian/special/110/>

* 1. 标题：//\*[@class="tit01"]
  2. 发布时间：
  3. 媒体名称：无
  4. 正文：//\*[@id="slidedesc2"]/p
  5. 评论获取方法：同新闻类

1. 政能亮：

例：http://pl.ifeng.com/opinion/zhengnengliang/241/1.shtml

* 1. 标题：normalize-space(//\*[@class="title"])
  2. 发布时间：normalize-space(//\*[@class="time"]) 日期格式：2017.06.08
  3. 媒体名称：凤凰网
  4. 正文：//\*[@class="nr clearfix"]/div/p
  5. 评论获取方法：同新闻类

1. 汽车类：

例：<http://auto.ifeng.com/haiwai/20170608/1089828.shtml>

* 1. 标题：//\*[@class="arl-cont"]/h3/span
  2. 发布时间：//\*[@id="pubtime\_baidu"] 日期格式：2017年06月08日 07:56:40
  3. 媒体名称：substring(//\*[@id="source\_baidu"],4)
  4. 正文：//\*[@class="arl-c-txt"]/p
  5. 评论获取方法：同新闻类

1. 房产类：

例：<http://dl.house.ifeng.com/detail/2017_06_09/51105916_0.shtml>

* 1. 标题：//\*[@class="title"]/h2
  2. 发布时间：//\*[@class="title"]/div[1]/span 日期格式：2017年06月09日 08:24
  3. 媒体名称：substring(//\*[@class="title"]/div[2]/span,4)
  4. 正文：//\*[@class="article"]/p
  5. 评论获取方法：同新闻类

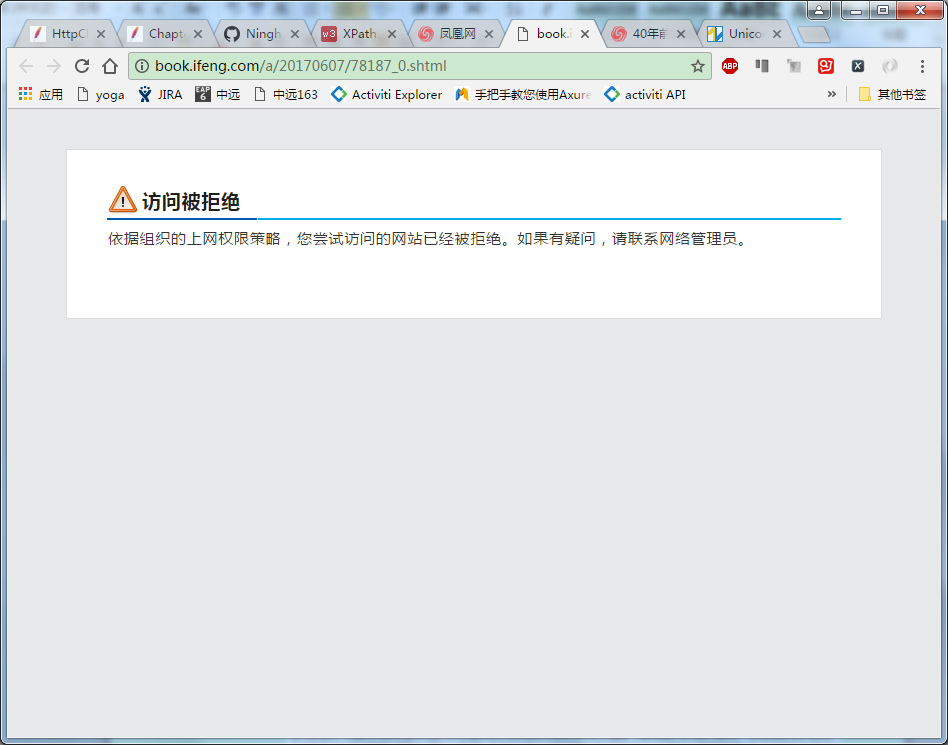
1. 青年类：

例：<http://young.ifeng.com/a/20170605/44630268_0.shtml>

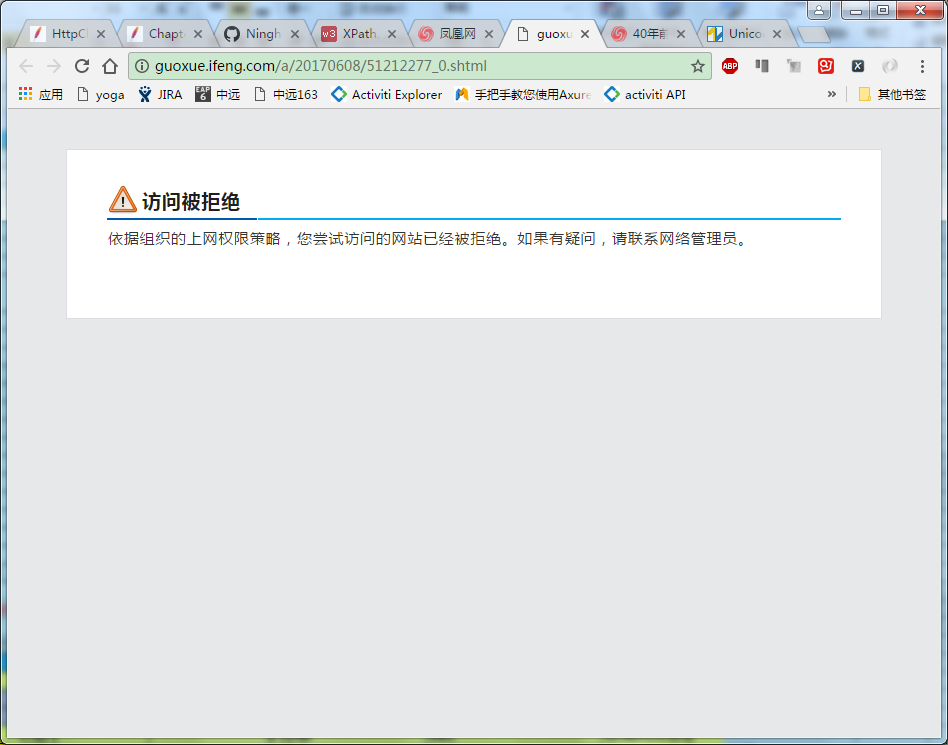
* 1. 标题：//\*[@class="ya\_main\_tit"]/h1
  2. 发布时间：//\*[@class="yamt\_time"]/span 日期格式：2017-06-05
  3. 媒体名称：substring(//\*[@class="yamt\_from"],4)
  4. 正文：//\*[@class="ya\_main\_con"]/p
  5. 评论获取方法：同新闻类

打不开网页：

* + - 1. book.ifeng.com开头的网页



* + - 1. guoxue.ifeng.com开头的网页



* + - 1. ([www.)yc.ifeng.com](http://www.)yc.ifeng.com)开头的网页

