**“泉港碳九泄漏事件”数据抓取需求**

抓取平台：新浪微博

搜索关键词：泉港 碳九

抓取日期：2018年11月4日 0时——2018年12月1日0时

抓取时段：每日18：00——次日0：00

抓取字段：博主id、博主名称、博主认证信息（个人认证or机构认证）、博主年龄、博主性别（这三个字段如果能抓下来更好）、博主主页、微博正文、博文独立网址、被转发的原微博、发布时间、转发数、评论数、点赞数

抓取需求：

1.需抓取“展开全文”后的全部博文内容

2.需抓取博文内容中包含所有的表情符号

3.因为18：00——次日0：00是发博的高峰时段，每小时的微博量都可能会达到50页。所以，中间有一些时段需按照微博搜索的最小时间单位——1个小时，来进行搜索和抓取，以便抓取到全部可显示的数据。

根据我已搜索的情况，搜索抓取的时段分别为：

|  |  |  |
| --- | --- | --- |
| 起点时间 | 终点时间 | 页数 |
| 2018.11.04 18:00 | 2018.11.05 00:00 | 2 |
|  |  |  |
| 2018.11.05 18:00 | 2018.11.06 00:00 | 5 |
|  |  |  |
| 2018.11.06 18:00 | 2018.11.06 19:00 | 50 |
| 2018.11.06 19:00 | 2018.11.06 20:00 | 50 |
| 2018.11.06 20:00 | 2018.11.06 21:00 | 50 |
| 2018.11.06 21:00 | 2018.11.06 22:00 | 50 |
| 2018.11.06 22:00 | 2018.11.06 23:00 | 50 |
| 2018.11.06 23:00 | 2018.11.07 00:00 | 50 |
|  |  |  |
| 2018.11.07 18:00 | 2018.11.07 19:00 | 50 |
| 2018.11.07 19:00 | 2018.11.07 20:00 | 50 |
| 2018.11.07 20:00 | 2018.11.07 21:00 | 28 |
| 2018.11.07 21:00 | 2018.11.07 22:00 | 33 |
| 2018.11.07 22:00 | 2018.11.07 23:00 | 29 |
| 2018.11.07 23:00 | 2018.11.08 00:00 | 50 |
|  |  |  |
| 2018.11.08 18:00 | 2018.11.08 19:00 | 50 |
| 2018.11.08 19:00 | 2018.11.08 20:00 | 50 |
| 2018.11.08 20:00 | 2018.11.08 21:00 | 50 |
| 2018.11.08 21:00 | 2018.11.08 22:00 | 50 |
| 2018.11.08 22:00 | 2018.11.08 23:00 | 50 |
| 2018.11.08 23:00 | 2018.11.09 00:00 | 50 |
|  |  |  |
| 2018.11.09 18:00 | 2018.11.09 21:00 | 43 |
| 2018.11.09 21:00 | 2018.11.10 00:00 | 46 |
|  |  |  |
| 2018.11.10 18:00 | 2018.11.10 23:00 | 48 |
| 2018.11.10 23:00 | 2018.11.11 00:00 | 9 |
|  |  |  |
| 2018.11.11 18:00 | 2018.11.12 00:00 | 26 |
|  |  |  |
| 2018.11.12 18:00 | 2018.11.12 22:00 | 49 |
| 2018.11.12 22:00 | 2018.11.13 00:00 | 19 |
|  |  |  |
| 2018.11.13 18:00 | 2018.11.14 00:00 | 22 |
|  |  |  |
| 2018.11.14 18:00 | 2018.11.15 00:00 | 35 |
|  |  |  |
| 2018.11.15 18:00 | 2018.11.16 00:00 | 5 |
|  |  |  |
| 2018.11.16 18:00 | 2018.11.17 00:00 | 3 |
|  |  |  |
| 2018.11.17 18:00 | 2018.11.18 00:00 | 4 |
|  |  |  |
| 2018.11.18 18:00 | 2018.11.19 00:00 | 1 |
|  |  |  |
| 2018.11.19 18:00 | 2018.11.20 00:00 | 11 |
|  |  |  |
| 2018.11.20 18:00 | 2018.11.21 00:00 | 9 |
|  |  |  |
| 2018.11.21 18:00 | 2018.11.22 00:00 | 2 |
|  |  |  |
| 2018.11.22 18:00 | 2018.11.23 00:00 | 2 |
|  |  |  |
| 2018.11.23 18:00 | 2018.11.24 00:00 | 1 |
|  |  |  |
| 2018.11.24 18:00 | 2018.11.25 00:00 | 1 |
|  |  |  |
| 2018.11.25 18:00 | 2018.11.25 19:00 | 18 |
| 2018.11.25 19:00 | 2018.11.25 20:00 | 42 |
| 2018.11.25 20:00 | 2018.11.25 21:00 | 44 |
| 2018.11.25 21:00 | 2018.11.25 22:00 | 50 |
| 2018.11.25 22:00 | 2018.11.25 23:00 | 50 |
| 2018.11.25 23:00 | 2018.11.26 00:00 | 50 |
|  |  |  |
| 2018.11.26 18:00 | 2018.11.27 00:00 | 9 |
|  |  |  |
| 2018.11.27 18:00 | 2018.11.28 00:00 | 9 |
|  |  |  |
| 2018.11.28 18:00 | 2018.11.29 00:00 | 3 |
|  |  |  |
| 2018.11.29 18:00 | 2018.11.30 00:00 | 1 |
|  |  |  |
| 2018.11.30 18:00 | 2018.12.01 00:00 | 2 |