

@[TOC](新闻爬虫及爬取结果查询网站的搭建(三))

## 概要

本节主要介绍对爬取结果进行查询的网站搭建。主要有以下几个部分：1、用pgsql查询已爬取的数据 2、用网页发送请求到后端查询 3、用express构建网站访问pgsql 4、用表格显示查询结果

- 对数据进行查询 其中用pgsql查询数据，在第一章pgSQL连接池部分已经介绍过，在这里构建sql查询语句并且调用pgsql.query\_noparam进行查询即可。

## 用express构建网站访问pgsql

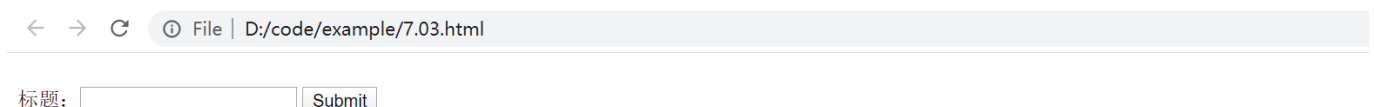
首先构造前端，代码如下：

```
<!DOCTYPE html>
<html>

<body>
  <form action="http://127.0.0.1:8080/process_get" method="GET">
    <br> 标题：<input type="text" name="title">
    <input type="submit" value="Submit">
  </form>
  <script>
  </script>
</body>

</html>
```

效果如下图：



[https://blog.csdn.net/yc\\_hong](https://blog.csdn.net/yc_hong)

后端获取查询的参数，并且在数据库中进行查询的代码如下：

```
var express = require('express');
var app = express();
var pgsql = require('./pg.js');
//app.use(express.static('public'));
app.get('/7.03.html', function(req, res) {
```

```
res.sendFile(__dirname + "/" + "7.03.html");
})

app.get('/process_get', function(req, res) {
  res.writeHead(200, { 'Content-Type': 'text/html;charset=utf-8' });
  console.log(req.query.title);
  var fetchSql = "select url,source_name,title,author,publish_date " +
    "from fetches where title like '%" + req.query.title + "%' ";
  pgsql.query_noparam(fetchSql, function(err, result, fields) {
    console.log(result.rows);
    res.end(JSON.stringify(result.rows));
  });
})

var server = app.listen(8080, function() {
  console.log("访问地址为 http://127.0.0.1:8080/7.03.html")
})
```

输入查询标题中包含“新冠”的新闻，得到的结果如下图：



# 用表格显示查询结果

在这里使用express脚手架来创建一个网站框架。在powershell中输入：

```
express -e Search_site
```

-e的参数表明我们用ejs作为缺省view引擎，而非采用jade 会生成search\_site文件夹，将之前写好的pg.js拷贝到该文件夹下。得到的文件目录如下图：

bin	2020/4/23 21:11	文件夹	
node_modules	2020/4/23 21:44	文件夹	
public	2020/4/23 21:22	文件夹	
routes	2020/4/23 21:11	文件夹	
views	2020/4/23 21:11	文件夹	
app.js	2020/4/23 21:11	JavaScript 文件	2 KB
package.json	2020/4/23 21:44	JSON 文件	1 KB
package-lock.json	2020/4/23 21:44	JSON 文件	20 KB
pg.js	2020/4/23 20:21	JavaScript 文件	2 KB

[https://blog.csdn.net/yc\\_hong](https://blog.csdn.net/yc_hong)

pg.js拷贝后还需要在search\_site文件夹内cmd运行：

```
npm install pg -save
```

--save表示将mysql包安装到该项目中，并且将依赖项保存进package.json里 在search\_site文件夹内cmd运行 npm install 将package.json中列举的依赖项全部安装，完成网站搭建。

对search\_site/routes/index.js进行修改，修改后的结果如下：

```
router.get('/process_get', function(request, response) {
  //sql字符串和参数
  console.log(request.query.title);
  console.log(request.query.sort);
  try {
    var fetchSql = "select url,source_name,title,author,publish_date " +
      "from fetches where title like '%" + request.query.title + "%' " +
      "order by " + request.query.sort;
    pgsql.query_noparam(fetchSql, function(err, result, fields) {
      if(err) {
        console.log(err)
      }
      response.writeHead(200, {
        "Content-Type": "application/json"
      });
      console.log(result.rows);
      response.write(JSON.stringify(result.rows));
      response.end();
    });
  } catch(e) {
    console.log(e);
  }
});
module.exports = router;
```

在search\_site/public/下创建一个search.html，在其中添加一个表格，并且为前面的提交按钮，绑定一个脚本。在点击时，清空id=record2的表格，然后添加表头，再对GET操作得到的每一行数据改成符合表格形式，最后添加到record2中。

```
<div class="cardLayout" style="margin: 10px 0px">
  <table width="100%" border="1" id="record2"></table>
</div>
<script>
  $(document).ready(function() {
    $("input:button").click(function() {
      var title = $("input:text").val();
      console.log("title = " + title);
      var params = '/process_get?title=' + $("input:text").val() ;
      $.get(params, function(data) {
        $("#record2").empty();
        $("#record2").append('<tr class="cardLayout"><td>url</td>
        <td>source_name</td>' +
          '<td>title</td><td>author</td><td>publish_date</td></tr>');
        for (let list of data) {
          let table = '<tr class="cardLayout"><td>';
          Object.values(list).forEach(element => {
            table += (element + '</td><td>');
          });
          table = table.substring(0, table.length - 4);
          $("#record2").append(table + '</tr>');
        }
      });
    });
  });
</script>
```

查询的到的结果显示如下：

url	source_name	title	author	publish_date
http://finance.eastmoney.com/a/202003291435717621.html	央视新闻	英国将境内新冠肺炎死亡病例数量控制在2万以下认作“好结果”	东方财富网	2020-03-28T16:00:00.000Z
http://finance.eastmoney.com/a/202003301436626261.html	中国日报网	美国FDA紧急批准医院利用氯喹和羟氯喹治疗新冠肺炎病人	东方财富网	2020-03-29T16:00:00.000Z
http://finance.eastmoney.com/a/202003301436238492.html	国家卫健委	国家卫健委：3月29日新增31例新冠肺炎确诊病例 其中境外输入30例	东方财富网	2020-03-29T16:00:00.000Z
http://finance.eastmoney.com/a/202003301436068966.html	新华网	约翰斯·霍普金斯大学：全球新冠肺炎确诊病例超过70万例	东方财富网	2020-03-29T16:00:00.000Z
http://finance.eastmoney.com/a/202003301436080993.html	人民日报	美国新冠病毒感染病例近14万	东方财富网	2020-03-29T16:00:00.000Z
http://finance.eastmoney.com/a/202003311438663039.html	新华社	《求是》发表习近平重要文章《在湖北省考察新冠肺炎疫情防控工作时的讲话》	东方财富网	2020-03-30T16:00:00.000Z
http://www.chinanews.com/cj/2020/03-31/9143067.shtml	中国新闻网	中国新冠肺炎AI诊断技术支援日本医院	房家梁	2020-03-30T16:00:00.000Z
http://finance.eastmoney.com/a/202003311438975945.html	环球网	外媒：全球新冠肺炎确诊病例超80万例	东方财富网	2020-03-30T16:00:00.000Z
http://finance.eastmoney.com/a/202003311438936671.html	海外网	西班牙新增9222例新冠肺炎确诊病例 累计94417例	东方财富网	2020-03-30T16:00:00.000Z
http://finance.eastmoney.com/a/202003311438947524.html	央视新闻客户端	伊朗新增新冠肺炎确诊病例3111例 累计确诊44606例	东方财富网	2020-03-30T16:00:00.000Z
http://finance.eastmoney.com/a/202003311438842850.html	央视新闻客户端	国家卫健委：4月1日起每日疫情通报中将增加新冠肺炎无症状感染者相关情况	东方财富网	2020-03-30T16:00:00.000Z
http://finance.eastmoney.com/a/202004011440011775.html	国家卫健委	国家卫健委：3月31日新增36例新冠肺炎确诊病例 新增无症状感染者130例	东方财富网	2020-03-31T16:00:00.000Z
http://www.chinanews.com/cj/2020/04-01/9144258.shtml	中国新闻网	三部门：4月起出口新冠病毒检测试剂等报关须提供声明	李玉素	2020-03-31T16:00:00.000Z
http://finance.eastmoney.com/a/202004031443816539.html	海外网	英国首相约翰逊持续出现新冠肺炎症状 将继续自我隔离	东方财富网	2020-04-03T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444187894.html	海外网	美国新冠肺炎累计确诊逾30万例 特朗普称将增派军队帮助抗疫	东方财富网	2020-04-04T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444241364.html	央视新闻	泰国新增102例新冠肺炎确诊病例 累计2169例	东方财富网	2020-04-04T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444212964.html	央视新闻	印度新冠肺炎确诊病例升至3374例 新增472例	东方财富网	2020-04-04T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444185282.html	央视新闻	西班牙首相：采取措施进一步遏制新冠病毒蔓延	东方财富网	2020-04-04T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444248967.html	海外网	法国新冠肺炎确诊病例新增7824例 累计90853例	东方财富网	2020-04-04T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444189665.html	国家卫健委	国家卫健委：4月4日新增30例新冠肺炎确诊病例 新增无症状感染者47例	东方财富网	2020-04-04T16:00:00.000Z
http://finance.eastmoney.com/a/202004051444196622.html	央视新闻	世卫组织：已有207个国家地区出现新冠肺炎病例	东方财富网	2020-04-04T16:00:00.000Z

## 对查询结果按字段排序

添加了一个对查询结果能够选择按照指定字段进行排序的功能。在前端添加了一个下拉菜单，默认按照日期进行排序

```
<div class="select" style = "float: left; margin-left: 40px;">
  <select class="form-control" name = "sort" id = "sortid" style = "float:
left;">
    <option value = "publish_date" selected>日期</option>
    <option value = "category">分类</option>
    <option value = "source">来源</option>
  </select>
</div>
```

然后后端的查询语句改成：

```
var fetchSql = "select url,source_name,title,author,publish_date " +
  "from fetches where title like '%" + request.query.title + "%'" +
  "order by " + request.query.sort;
```

## 对查询结果进行分页

在这部分，使用了bootstrap table插件，使用了其提供的分页和其他功能。在使用时可以参考[bootstrap table 官方文档](#) 在使用HTML5时应该在其中添加：

```
<!doctype html>
<html lang="en">
  <head>
    <!-- Required meta tags -->
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-
fit=no">
    <title>Hello, Bootstrap Table!</title>

    <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.2.1/css/bootstrap.min.css"
integrity="sha384-
GJzZqFGwb1QTTN6wy59ffF1BuGJpLSa9DkKMP0DgiMDm4iYMj70gZWKYbI706tWS"
crossorigin="anonymous">
    <link rel="stylesheet"
href="https://use.fontawesome.com/releases/v5.6.3/css/all.css" integrity="sha384-
UHRtZLI+pbxtHCWp1t77Bi1L4ZtiqrqD80Kn4Z8NTSRyMA2Fd33n5dQ8lWUE00s/"
crossorigin="anonymous">
    <link rel="stylesheet" href="https://unpkg.com/bootstrap-
table@1.15.3/dist/bootstrap-table.min.css">
  </head>
  <body>
    <script src="https://code.jquery.com/jquery-3.3.1.min.js" integrity="sha256-
FgpCb/KJQlLNfOu91ta32o/NMZxltwRo8QtmkMRdAu8=" crossorigin="anonymous"></script>
    <script
src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.14.6/umd/popper.min.js"
integrity="sha384-
```

```
wHAIffRlMFy6i5SRaxvf0CifBUQy1xHdJ/yoi7FRNXMRBu5WHdZYu1hA6Z0blgut"
crossorigin="anonymous"></script>
<script
src="https://stackpath.bootstrapcdn.com/bootstrap/4.2.1/js/bootstrap.min.js"
integrity="sha384-
B0UglyR+jN6CkvvICOB2joaf5I4l3gm9GU6Hc1og6Ls7i6U/mkkaduKaBhlAXv9k"
crossorigin="anonymous"></script>
<script src="https://unpkg.com/bootstrap-table@1.15.3/dist/bootstrap-
table.min.js"></script>
</body>
</html>
```

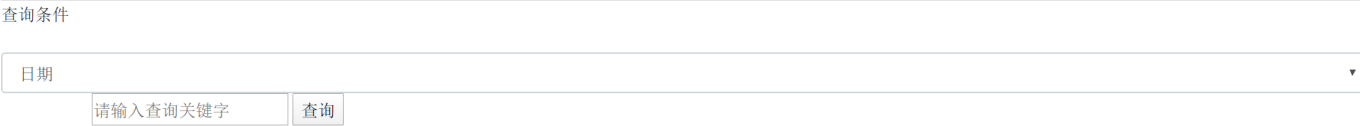
- 使用 在这里我使用的是通过javascript来以表格的形式显示数据。通过GET方法访问给定的url，并且将返回的json文件，解析生成表格，注意下面field的值应该和返回的json文件对应属性的keyvalue相同。生成表格的代码如下：

```
$(function(){
    console.log(params);
    $("#record2").empty();
    $('#record2').bootstrapTable({
        url:params,
        method:'GET',
        pagination:true,
        sidePagination:'client',
        pageSize:5,
        striped : true,
        sortable : true,
        sortOrder:"asc",
        showRefresh:true,
        search:true,
        showToggle: true,
        toolbar: '#toolbar',
        showColumns : true,
        columns:[{
            field : 'url',
            title : 'url'
        }, {
            field:'title',
            title:'title'
        }, {
            field:'source_name',
            title:'source_name'
        },{
            field:'author',
            title:'author'
        },{
            field:'publish_date',
            title:'publish_date'
        }
    ]
})
```

```
    })
  });
}
```

## 前端界面

在前面设计得到的界面，没有调整格式的效果是下图这样的：



这个界面非常丑，下面将逐一对其进行修改。

- 对查询条件那一行进行字体大小的更改和字体更改 代码如下：

```
<style>
  .title {
    border:1px solid rgb(199, 193, 193);
    font-family : 'FangSong';
    font-size : x-large;
  }
</style>
<div class = "title" >
  新闻查询
</div>
```

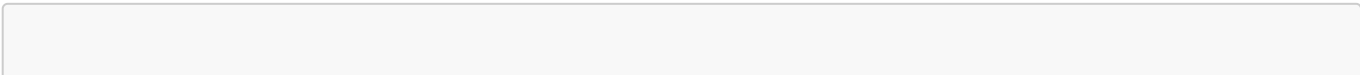
- 改变查询按钮格式

```
<input type="button" class="form-submit" value="查询">
```

对应设置样式的css：

```
.form-submit {
  background-color: rgb(0, 75, 214);
  border-radius:5px;
  color: #fff;
  font-size: 20px;
  border: none;
}
```

- 设置复选框、按钮、输入框在同一行 为了美观，将查询条件的设置放在一行。从上面可以看到输入框和提交按钮在同一个div中，下拉菜单在另一个菜单中。为了使他们在同一行，使用了float和inline显示进行调整。代码如下：





```
.select {
  display: inline;
}
.container {
  display: inline;
}
```

```
<div class="select" style="float: left; margin-left: 40px;">
  <select class="form-control" name="sort" id="sortid" style="float: left;
">
    <option value="publish_date" selected>日期</option>
    <option value="category">分类</option>
    <option value="source">来源</option>
  </select>
</div>
<div class="container" style="float: left; width: 100%;">
  <form action="" class="parent">
    <input type="text" class="search" name="title_text" placeholder="请输入查询
关键字" style="border-radius: 8px; ">
    <input type="button" class="form-submit" style="float: right" value="查
询">
  </form>
</div>
```

- 添加边框和背景 为了使页面风格同一且标题栏和查询栏有分割，对标题和查询条件部分添加了边框和背景色。代码如下,其中class = parent是查询条件部分所在的div：

```
.parent {
  border: 1px solid rgb(199, 193, 193);
  background: #ffffff;
}
/*标题*/
.title {
  border: 1px solid rgb(199, 193, 193);
  font-family: 'FangSong';
  font-size: x-large;
}
```

## 展示

下面是最终的输入页面

新闻查询

日期 ▾

请输入查询关键字

查询



输入股票进行查询得到的结果如下图：

查询条件

日期 ▾

股票

查询

Search

↻

🔍

⌵

url	title	source_name	author	publish_date
http://finance.eastmoney.com/a/202004071446069148.html	特高压建设高峰将至 机构给出股票池！逾6亿元大单火速布局这11股！	证券日报网	东方财富网	2020-04-07T16:00:00.000Z
http://finance.eastmoney.com/a/202004101449328119.html	股票基金再现一日售罄！认购金额超上限80亿	每日经济新闻	东方财富网	2020-04-09T16:00:00.000Z
https://xueqiu.com/1184824257/146486791	第一次！中央重磅发文，完善股票市场基础制度！发行、交易、退市、分红都有涉及，来看十大关键定调	雪球网	券商中国	2020-04-09T16:00:00.000Z
https://xueqiu.com/7228103229/146637028	格力电器年报前，只有一天可以回购股票	雪球网	股灾亲历者	2020-04-11T16:00:00.000Z
https://xueqiu.com/6565545060/147047174	清仓了房地产股票，苦海无边回头是岸	雪球网	价投傻瓜	2020-04-15T16:00:00.000Z

Showing 16 to 20 of 24 rows

5 ▴

rows per page

⏪

1

2

3

4

5

⏩

https://blog.csdn.net/ye\_hong