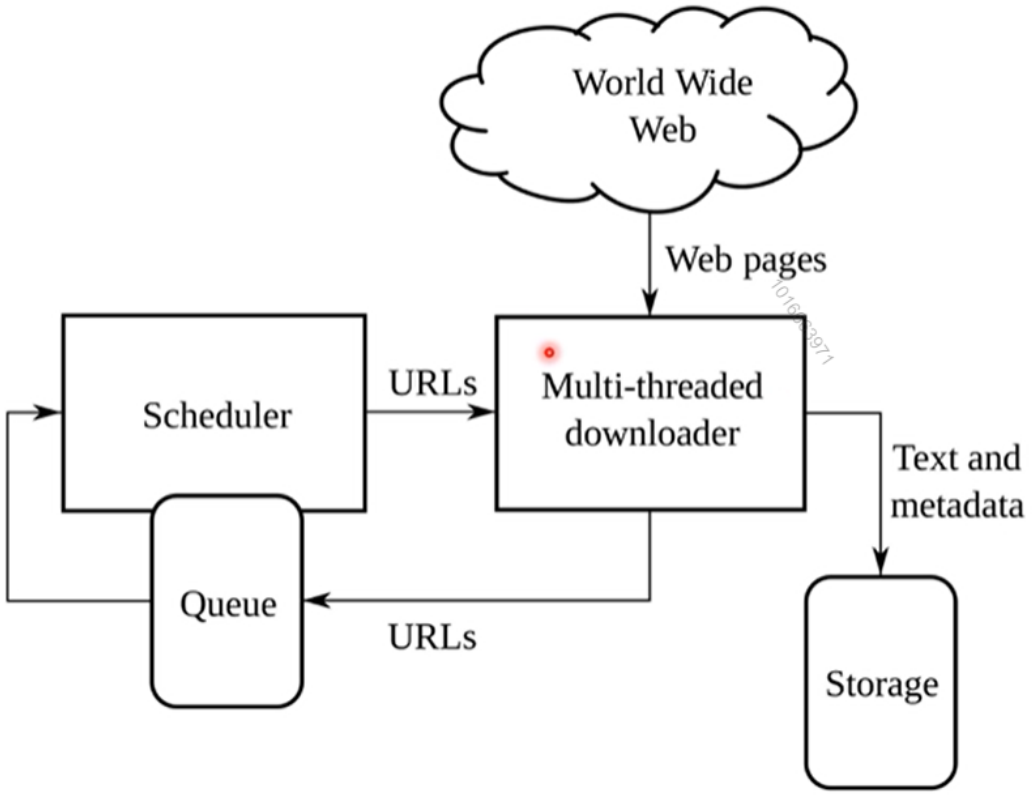
课程目标

爬取B站数据

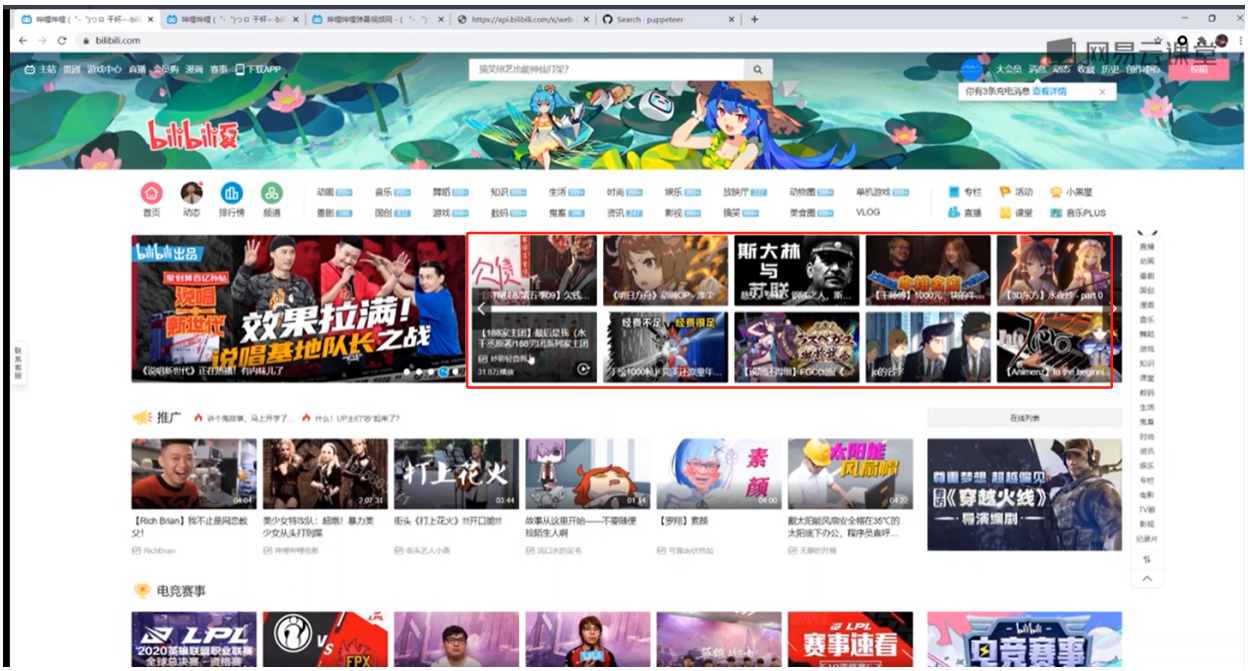
基本概念：



scheduler(一个算法，他的作用是调度)，有的时候我们不只使用一个爬虫，比如你要爬互联网上的数据，互联网上有多少的数据呢？据这种不太专业的数据统计，也有上亿个网站，要把这些网站都爬掉的话都需要调度的，即使你不像百度那样需要爬所有的网站，如果需要爬B站的数据，也是很多的。所有需要一个调度器的，通常爬虫的话，一台机器还不够，可能需要很多机器。 爬虫也可以做水军，模拟登录，很多评论都是由爬虫弄进去的，并不是用户这真实的评论。scheduler找到urls，发送给downloader下载器，他的主要作用是把

数据下载下来，然后把数据存到Storage里面去。下载器downloader下载后，会进行分析，然后产生新的URLs，比如今天爬B站，我们从B站的首页拿到很多卡片，卡片点进去的需要一个新的连接，跳转。

实战(B站有价值就是相关卡片数据，然后相关详情里面的相关推荐)



node服务端爬部分公开api

```
const fetch = require('node-fetch')
const app = require('express')()

function fetchData(){
 return new Promise((resolve,reject)=>{
 fetch('https://api.bilibili.com/x/web-interface/ranking/region?rid=1&day=3&original=0')
  .then(res=>{
 return res.json()
 }).then(res=>{
  resolve(res)
  })
  })
}
app.get('/',(req,res)=>{
 async function BiData(){
 let data = await fetchData()
 res.json({stats:200,data,message:'成功'})
```

```
  }
  BiData()
})


app.listen(9999,()=>{
  console.log('9999服务');
})
```

爬虫结果：

{"stats":200,"data":{"code":0,"message":"0","ttl":1,"data":[{"aid":"247837944","bvid":"BV1Pv411j79y","typename":"短片·手书·配音","title":"【番嘎之交】2022年春晚小品","subtitle":"","play":3585765,"review":3934,"video_review":4390,"favorites":59388,"mid":67527408,"author":"胡基霖","description":"表演者：番子，嘎子，胡子\n故事梗概：生活中有很多小商小贩缺斤短两，这不今天番子和嘎子就因为谁发生了瞎哭皆非的故事，嚯，他们来啦~","create":"2021-04-25 11:36","pic":"http://i0.hdslb.com/bfs/archive/fd4625cd58c837cff18efca3a20e65a7facfb18e.jpg","coins":177721,"duration":"3:30","badgepay":false,"pts":1948188,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"587864408","bvid":"BV1UB4y1c7SU","typename":"短片·手书·配音","title":"梅花十一：我只会心疼哥哥","subtitle":"","play":1563720,"review":2048,"video_review":2971,"favorites":28593,"mid":20555379,"author":"紫光-君","description":"整活类手书，仅供娱乐，看到最后有惊喜\n\n谢谢大家的喜欢，希望看到最后，再顺便帮up点个赞","create":"2021-04-26 11:58","pic":"http://i1.hdslb.com/bfs/archive/8e6237919fde8623cd2828e12b9573bda45ef0c9.jpg","coins":42296,"duration":"1:41","badgepay":false,"pts":1735633,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"375365232","bvid":"BV1rZ4y1F7tV","typename":"综合","title":"生 儿 为 人 我 很 抱歉","subtitle":"","play":1154052,"review":4027,"video_review":3714,"favorites":28516,"mid":321422126,"author":"进击的金厂长","description":"像我这样的人，如何过完这一生…","create":"2021-04-25 20:50","pic":"http://i1.hdslb.com/bfs/archive/81afae7ef358e6600522a1b905de7ae2926cb8e9.jpg","coins":110491,"duration":"5:56","badgepay":false,"pts":1246324,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"845299724","bvid":"BV1n54y1j7ro","typename":"综合","title":"派蒙只会心疼旅行者哥哥（多多配音版）","subtitle":"","play":815595,"review":1843,"video_review":688,"favorites":17684,"mid":534963525,"author":"多多poi二创组","description":"原视频：BV1fA41157ts（感谢up主授权）感谢多多的献声!!!!","create":"2021-04-26 12:00","pic":"http://i2.hdslb.com/bfs/archive/bde035d1b12c8885318d1444dd195dae6d41624a.jpg","coins":19161,"duration":"0:34","badgepay":false,"pts":1059560,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":1,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"205242492","bvid":"BV1sh411S767","typename":"MAD·AMV","title":"五分钟带你回顾12年来血洗B站的动漫神曲","subtitle":"","play":647542,"review":1890,"video_review":2961,"favorites":39841,"mid":293556783,"author":"某科学的未来预测","description":"UP主：某科学的未来预测\n视频制作：某科学的超万能王\n字幕：洛天依___\n后期：辽之岛绫小路，蔷薇_原澪丶sab\n审核菌辛苦了，所有创作团队成员积极参与视频创作，求一次过审，要不然我还要拿着截图来稿件申诉，好麻烦的\n\n声明：本视频为二次创作\n策划/演唱：暮肆_Mousse\n编曲/吉他/贝斯/架子鼓/钢琴/乐器混音：张小旭\n人声混音/母带：圈太Studio\n\n串烧歌曲歌单：\n2009-《only my railgun》\n2010-《恋愛サーキュレーション》\n2011-《secret base~君がくれたもの》","create":"2021-04-24 10:00","pic":"http://i0.hdslb.com/bfs/archive/2631ba633c2eacfefacd41301cb5d286e654dec3.jpg","coins":52354,"duration":"5:12","badgepay":false,"pts":1023941,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":0,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":1,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"460257300","bvid":"BV1A5411w7ri","typename":"MAD·AMV","title":"【查古风】喵，走！","subtitle":"","play":326740,"review":744,"video_review":550,"favorites":24994,"mid":11255948,"author":"Baka恶魔","description":"高技术力向合作~创意转场~\nBGM：Break My Heart -Dua Lipa\n素材：辉夜大小姐想让我告白 物语系列 国家队 等\n求推荐！！\n封面：baka\n分段情况评论区置顶","create":"2021-04-28 11:30","pic":"http://i1.hdslb.com/bfs/archive/af55eebb20d3abb23ffa467f3a21231c0576e8f1.jpg","coins":21319,"duration":"2:13","badgepay":false,"pts":979579,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":0,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":1,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"672722743","bvid":"BV1aU4y1b7hn","typename":"综合","title":"神曲一响，万人跟唱","subtitle":"","play":1601883,"review":1232,"video_review":1378,"favorites":35080,"mid":130130236,"author":"枫泽雨畔","description":"BGM:《唱脸谱》——杭天琪\n初中音乐课中最洗脑歌曲，挥之不去。喜欢的给个三连吧，求求了。\n《老鼠出售》《寂寞的老鼠》\n《请保持安静》《西部牛仔》\n《汤姆猫演唱会》《博士教学班》\n《海底新世界》《荒岛假期》\n《斑点猫》《小孤儿》\n《祖鲁猫》《猫咪宝贝》\n《秃鹰变奏曲》《老鼠舞厅》\n《老鼠打扫记》《有理的宝贝》\n《可怕的白鼠》《汤姆约会记》\n《耶诞夜》《拿坡里老鼠》\n《天上普斯》《捕鼠大队》\n《小海豹失踪记》《周末夜之猫》","create":"2021-04-24 12:00","pic":"http://i0.hdslb.com/bfs/archive/b1894b7b23775f4406fed96c77aaf12ebf2d5471.jpg","coins":66060,"duration":"1:46","badgepay":false,"pts":978749,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"845356042","bvid":"BV1254y1j7nR","typename":"综合","title":"一天30根烟，他终于得了肺癌","subtitle":"","play":523431,"review":1486,"video_review":1778,"favorites":7016,"mid":7487399,"author":"努力的Lorre","description":"Hellblazer #41-46 \nDangerous Habits 致命嗜好","create":"2021-04-28 19:05","pic":"http://i0.hdslb.com/bfs/archive/bc35e8c2f58729d288288c2721f82b66493a14a7.jpg","coins":15378,"duration":"17:43","badgepay":false,"pts":899484,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":0,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"460171055","bvid":"BV1m5411c7sP","typename":"MAD·AMV","title":"惊艳了无数人的完美圈面，你认识多少呢？","subtitle":"","play":671324,"review":579,"video_review":787,"favorites":42482,"mid":10773628,"author":"永远的我的马宠Me","description":"每一帧都是壁纸系列~\n无水印可拿去做壁纸使用，只要不盗用就好。\n制作不易，还请三连支持一下，阿里嘎多~\nBGM: Martin Ermen - River Flows In You","create":"2021-04-23 17:35","pic":"http://i2.hdslb.com/bfs/archive/78e645992d7535bf977fbf12843c7e00779041ab.jpg","coins":42549,"duration":"2:28","badgepay":false,"pts":822014,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"760300608","bvid":"BV1R64y1y7mX","typename":"短片·手书·配音","title":"《 艺 术 不 可 以 搬 地 府》!","subtitle":"","play":352002,"review":806,"video_review":2143,"favorites":4873,"mid":90361813,"author":"三十六贱笑","description":"本期的内容其实夹杂了很多自己的情感输出\n做了自己想做的反转世界，真实是什么，每个人心中都有自己答案。\n真的要感谢汤达人给我这样任性的机会，完成了一期看起来不像商单的商单！\n感谢体谅我的任性！最后还是希望大家看的开心！你的三连是我最大的动力，我们下期见！","create":"2021-04-28 11:30","pic":"http://i2.hdslb.com/bfs/archive/dfa16e4914fe3e167192dced6fe2314ad3cf10cf.jpg","coins":16982,"duration":"2:47","badgepay":false,"pts":779778,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":1,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":0,"ugc_pay_preview":0,"no_background":0},"redirect_url":""},
{"aid":"587813346","bvid":"BV1sB4y1c7Px","typename":"短片·手书·配音","title":"【配音小剧场】罗德岛幼儿园","subtitle":"","play":262150,"review":803,"video_review":484,"favorites":8806,"mid":11742550,"author":"哈米伦的弄笛者","description":"这个罗德岛幼儿园刚看到就觉得超棒，就拉着一群朋友做了个配音，大家都配的特别认真！如果可以的话我想做个幼儿园小剧场2333\n感谢配音 博士：魔法Zc目录\n 飘尔希：陶二梦\n 能天使：小虾鱼Official\n 霜叶：田汐汐_Official\n 杰西卡：蓬花泡泡\n剪辑：极冰银月 筠箬piano","create":"2021-04-28 11:30","pic":"http://i2.hdslb.com/bfs/archive/5abcbe51f915048874da658bcd0814f4cefe6d5d.jpg","coins":9650,"duration":"1:43","badgepay":false,"pts":650723,"rights":{"bp":0,"elec":0,"download":0,"movie":0,"pay":0,"hd5":0,"no_reprint":1,"autoplay":1,"ugc_pay":0,"is_cooperation":1,"ugc_pay_preview":0,"no_background":0},"redirect_url":""}]},"message":"成功"}

但是有些需要登录的api就需要模拟登录了

```
{"stats":200,"data":{"code":-101,"message":"账号未登录","ttl":1},"message":"成功"}
```

模拟登录

手动添加上token再去请求数据。

```
const fetch = require('node-fetch')
```

```
const app = require('express')()

function fetchData(){
 return new Promise((resolve,reject)=>{
 fetch('https://api.bilibili.com/x/member/web/account',{
 headers:{'cookie':'_uuid=00B3D359-2310-1DA1-5D34-E2161E2D27B670006infoc; buvid3=987581A0-8085-4350-9271-FCF22039A337185014infoc; bsource=search_baidu; buvid_fp=987581A0-8085-4350-9271-FCF22039A337185014infoc; SESSDATA=c365f90b%2C1635237967%2Ce5e94%2A41; bili_jct=bcf74cb7f953cbdcbd80b25eb5d6a6f9; DedeUserID=17868069; DedeUserID__ckMd5=cb00c0c926e3d66e; sid=a9g9b6qs; bfe_id=6f2695e1895fb89897286b11ddc486b0; bp_video_offset_17868069=517818518683541265; fingerprint3=a318682c183fc382c274fb22e5789b9f; fingerprint=336eb44fea45f179a053172b33f62e24; fingerprint_s=590ece7250adf7287ffc8227dda15e48; buvid_fp_plain=987581A0-8085-4350-9271-FCF22039A337185014infoc; PVID=2'}
 })
 .then(res=>{
 return res.json()
 }).then(res=>{
 resolve(res)
 })
 })
}
app.get('/',(req,res)=>{
 async function BiData(){
 let data = await fetchData()
 res.json({stats:200,data,message:'成功'})
 }
 BiData()
})


app.listen(9999,()=>{
 console.log('9999服务');
})
```

上面说的这些都是固态话，要更加浏览器化，得需要重新像一个浏览器一样。需要一个浏览器内核这时候可以再github找到一个库Puppeteer谷歌内核一个库