新浪微博数据抓取解析

# 在选择抓取方式上面，我们选择Selenium（ThoughtWorks公司开发的web自动化测试工具） 开发工具为VS2013 。

# 微博有很好的防爬虫机制， 使用常用的解析网页会遇到无法获取所需元素（例如微博的评论以及转发的详细信息）。而 Selenium可以做到模拟用户正常浏览过程中的一系列操作，这样我们可以获取模拟操作后的页面的所有元素，

# 抓取过程：

# 首先模拟打开浏览器（这里我们选择火狐浏览器，火狐浏览器与Selenium有很好的兼容性），并跳转到话题所在的微博详情页。

# 这时我们可以通过Selenium获取到我们可以看到的页面上的所有信息，通过获取DOM元素，定位到该条微博的转发以及评论所在的地方。

# 之后将我们爬虫得到的数据暂时保存，根据我们的需求，将数据分类并处理。

# 接下来将处理好的数据保存到本地的数据库中，以供以后使用。