探索微博热门事件—报告

林丽静 厉扬豪(1100012740) 聂奕凝

1. **总览**

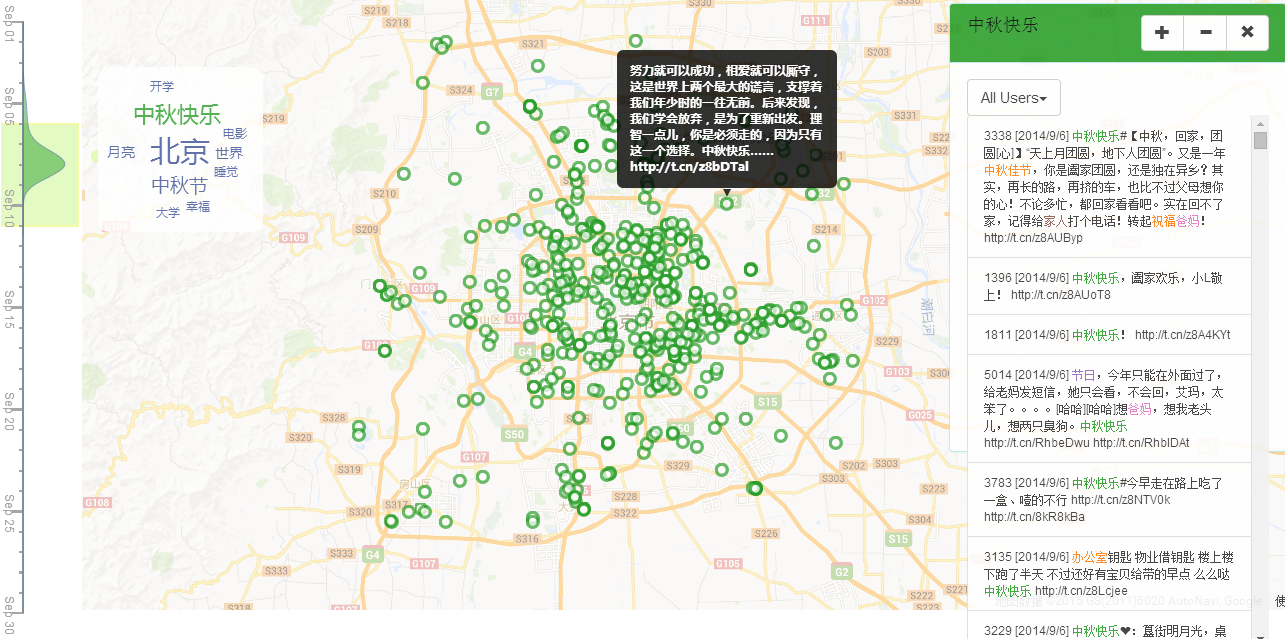


图1、 整体效果，左侧河流表示包含关键词“中秋快乐”的微博量随时间变化的趋势；地图上每一个圆表示包含关键词“中秋快乐”的微博；右侧浮动面板，展示了详细的相关微博内容。

这个作品支持以下四种探索：

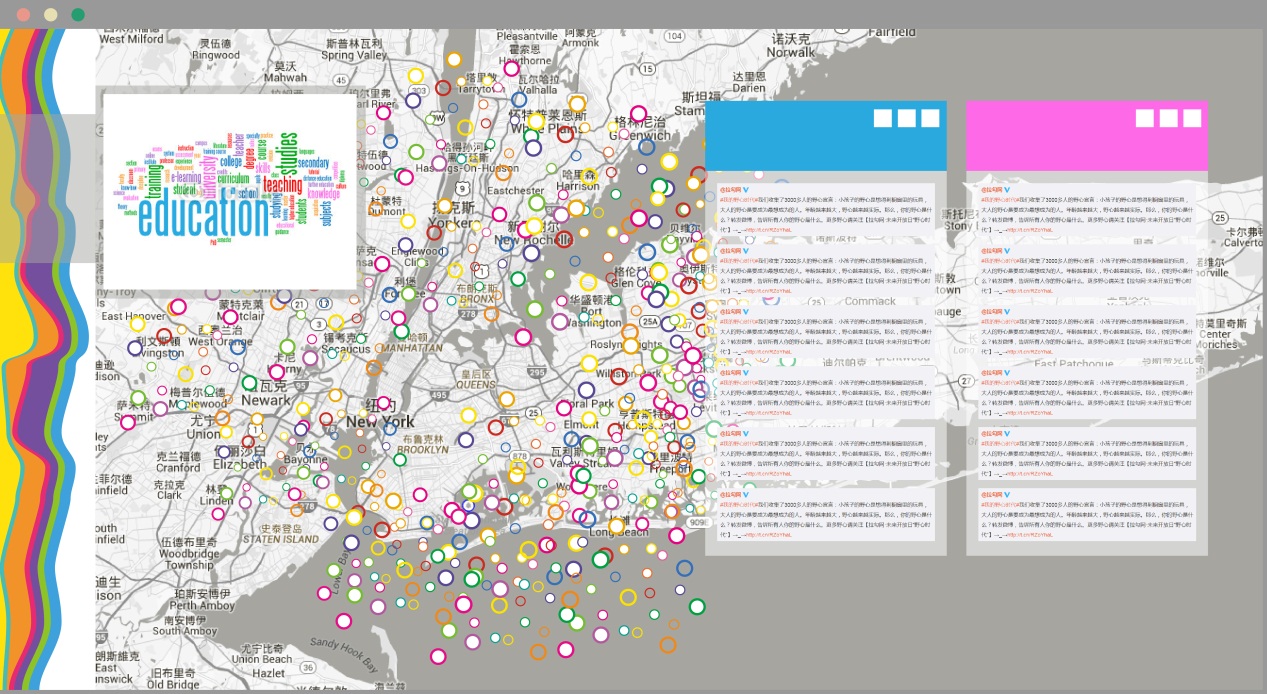
1. 探索任意一段时间内，微博热门事件；
2. 探索热门事件的地理分布特征；
3. 探索热门事件之间的内在联系；
4. 协助使用者发现广告用户。
5. **数据分析与处理**

原始数据给出了144191条时间范围为(2014.9.1到2014.10.10)在北京地区的带地理标签的微博。每条微博给出了其发布的用户、时间、地点以及微博的内容。根据我们设计的想法，考虑分析微博内容的关键词，从而发现热门事件。

关键词的提取我们采用了NLPIR汉语分词系统，对所有的微博内容进行了关键词提取与统计。在后期的使用中，对于一段时间的关键词权重的定义，考虑了1. 软件返回的关键词权重; 2. 这段时间内关键词出现的次数；3. 关键词在一段时间出现的次数占总时间的比例。

有了关键词之后，再通过对数据整理，获得关于各条微博，各个用户以及各个关键词对应的信息，处理完的数据存储成json格式，供后续可视化的实现。

1. **设计**

****

上图就是一开始对layout的一个设计图。整个应用包括地图、流图、标签云、微博浮动框、点线图几个主要视图组成。

展示地理信息的可视化最好为地图信息留出足够的空间，因此基础页面只有地图和流图两个视图，伴随后续操作逐步展开其他视图。同时，出于引导用户操作流程和界面整洁的原因，微博浮动框等几个视图都可以折叠和移动。

使用颜色进行信息分类是很直观的办法，但面对较多信息的时候色彩很容易混淆，这是后续可以考虑和改进的地方。

1. **可视化结果**

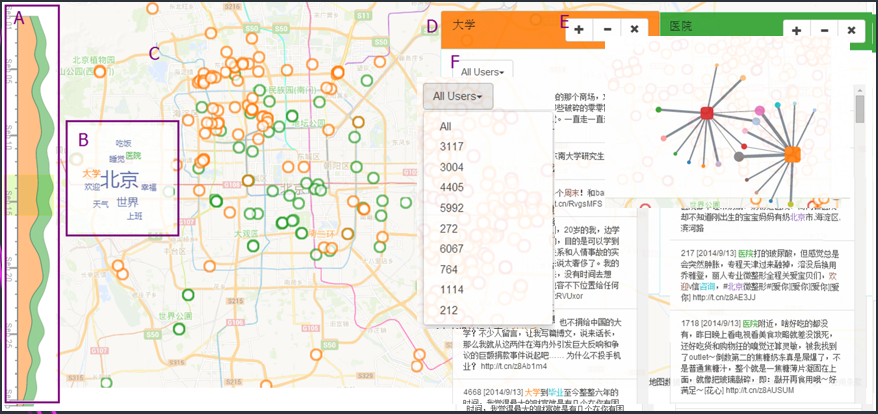


图2、 功能分析图

**A：**主题河流。展示关键词随时间的变化趋势；

**B：**标签云。展示感兴趣时间段内的关键词；

**C：**微博展示图。地图上每一个圆表示一条微博。从地图上，可观察与关键词相关的微博的地理分布特征。

**D：**微博浮动框。使用者可在此看微博的具体内容，此处微博按时间排序。同时，用户hover某一条微博，C（微博展示图）中相应的圆将变大，方便使用者查看任一条微博的地理位置。

**E：**点边图。图中方形表示关键词，圆表示users。通过这个图，可以观察任意多个关键词的内在联系。如果关键词的大部分主要users是共同拥有的，那么我们应该怀疑这些关键词的内在联系。

**F：**users列表。当使用者在E中发现感兴趣的用户时，可以在这里选择该用户，C（微博展示图）将展示出该用户在该时间段内发送的微博。同时，**D：**（微博浮动框）内的微博将该用户的微博置前。

1. **案例分析**
2. **9月8日，中秋节**

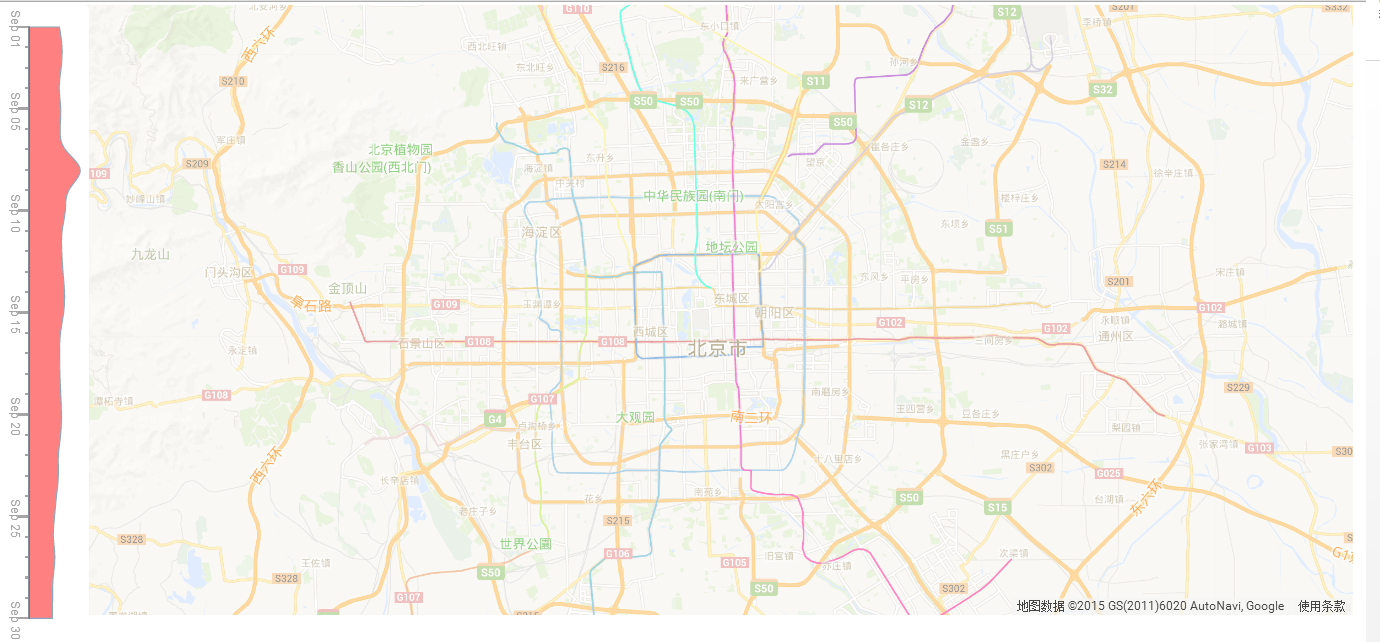


图3、初始化图

加载结束之后，主题河流显示9月1号到9月30号期间，微博随时间的变化趋势。9月8号左右，微博量突然激增。选中这段时间，观察这段时间内的热门事件。



图4、选择9月7号到9月9号时，展示的主要关键词

通过关键词云，可以发现应该是“中秋节”引起微博量的激增。



图5、“中秋节”与“中秋快乐”

选中这段时间“中秋节”和“中秋快乐”，查看两个关键词top users之间的联系。发现十个top users里面，有5个是两个关键词共同拥有。如此紧密的联系，表明这两个关键词相关性很大。显然，“中秋快乐”与“中秋节”说的就是同一个事。这个例子，可以说明，我们工具的有效性。当数据存在关系紧密的关键词，它可以协助使用者发现他们。

**2、 张杰9月20日演唱会**

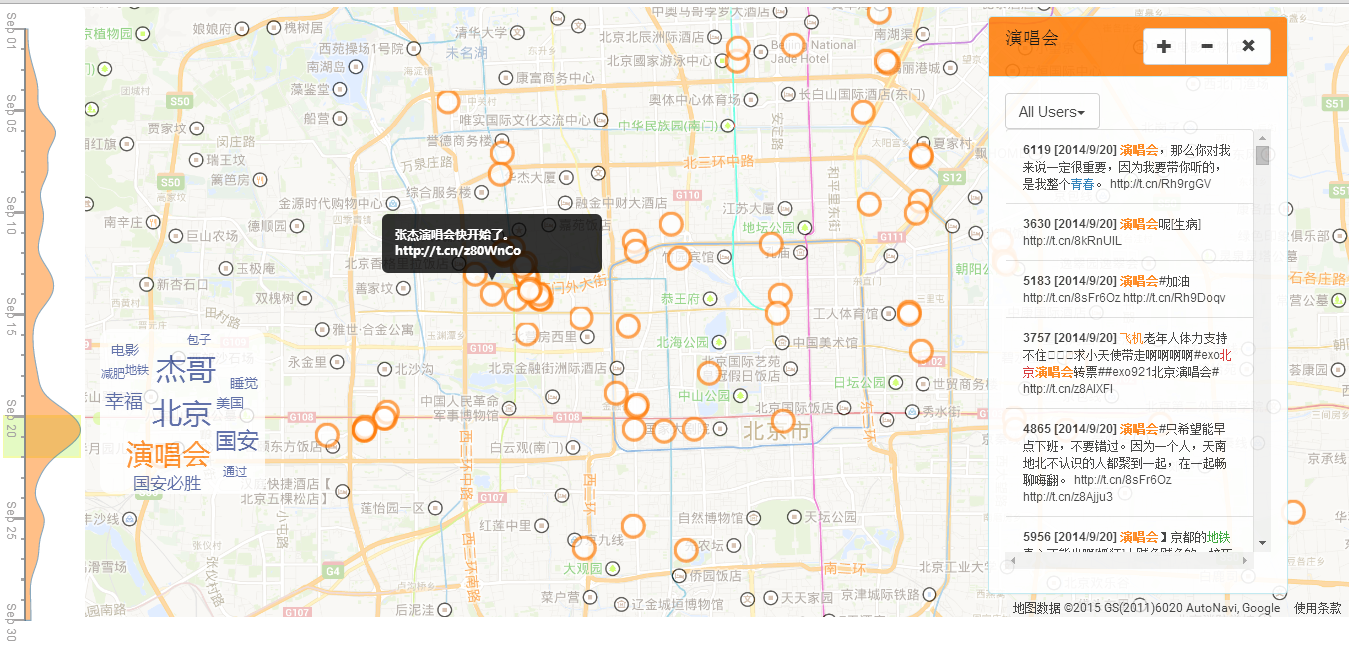


图5、探索“演唱会”图

选择9月20附件这段时间，选择“演唱会”。观察主题河流，发现北京9月份，一直都有与“演唱会”相关的微博。但是9月20左右的微博量很大。关键词云上大大的“杰哥”，已经提示我们这可能与张杰的演唱会相关。查看微博列表，可以发现，这个高峰确实是由杰哥的演唱会引起的。通过微博在地图上的展示，可以发现首都体育馆附件的微博特别的多，可以怀疑该演唱会就是在此举办的。鼠标hover在这些点上，可以发现，张杰的演唱会是9月20号在首都体育馆举办的。

1. **总结**
2. 关键词云显示的关键词个数是程序设定的，应该提示工具，让用户选择关键词个数；
3. 删除地图上的图表。虽然适度减小了地图的亮度，但是当将地图进行放大时，地图上的图标会影响视图。
4. 关键词-users关系图，没有添加有效的交互功能。理论上，可以在users上，添加交互功能，方便用户选择感兴趣的用户，甚至可以直接链接到该用户的主页。但是，这个涉及到用户的隐私，而且，相应的链接并不好获取。