

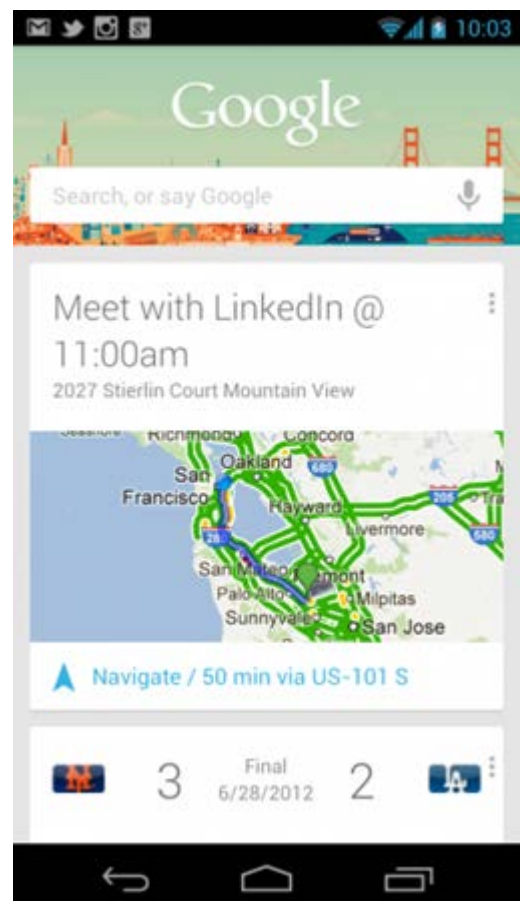
# 基于移动设备的用户行为记录分析平台

## 任务简介

### 系统功能需求

本次大作业需要大家利用学习到的 java 知识和方法来构建一个实际的应用（系统），题为“基于移动设备的用户行为记录分析平台”，将要实现以下核心功能：

- **利用手机收集大家一天的行为和状态数据**，包括：
  - 个人数据：位置（GPS，WIFI）、当前状态（图片、语音、文字描述）、运动状态（静止，行动等）；
  - 环境数据：时间戳、温度、湿度、光强等。
- **根据采集的数据，在手机客户端提供一些有效的提示信息来服务用户**（类似 Siri 和 Google now 的生活助手）：
  - 健康提示：定时喝水，定时活动；
  - 事件提示：根据天气提示出行，根据日历提示先关事宜；
  - 根据采集的通话记录给予一部分消费提示（隐私设置）
  - 社交需求：在客户端上分享自己的“数据”到微信，一种深度的状态发布。
- **用户行为数据的深度挖掘与展示**：
  - 手机客户端采集的数据上传到服务器端；
  - 利用 PC 端强大的处理能力和丰富的展示形式来对数据进行分析 and 展示。
  - 利用时间的主维度，可以对自己的“轨迹”进行管理。
  - 分析并计算一些间接的信息，并可以通过社交网站分享（新浪微博）。



## 基本要求

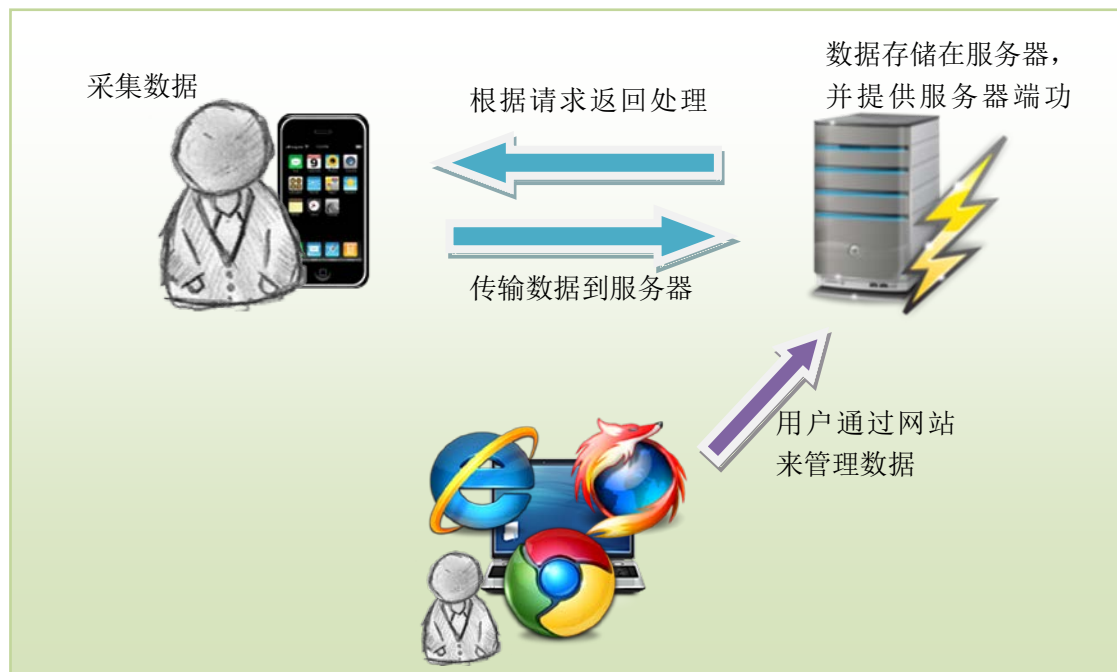
本次大作业由客户端应用和服务器端两个主要模块组成。每个模块各为 100 分满分，学生可以任选其一完成。鉴于客户端应用需要采集很多本地数据，作业要求实现客户端的学生需要配备真实的安卓手机用来最后的演示。

由于大作业涉及到客户端和服务器端的通信，学生需要组队（两个人一组）分别实现客户端和服务器端。最后大作业检查的时候也要求一组的学生一同来配合演示。但作业最后的评分只检查学生自己负责实现的那一模块，即客户端应用和服务器端分别评分互不影响。

下面将详细讲述本次大作业功能实现要求。

## 系统体系结构描述

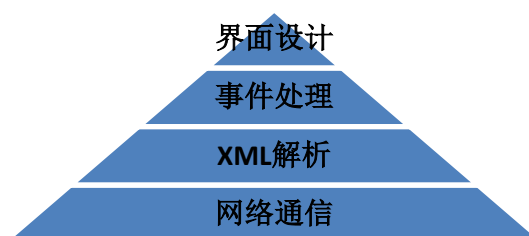
本次大作业需要开发的系统结构如下图所示，首先我们需要开发 **Android** 客户端采集用户的感知数据，并且能够与服务器端进行通信，将数据保存在服务器端。另外，本次作业还需要开发一套网站来帮助用户管理存放在服务器上的数据。各个部分的关联大致如下：



下面将详细讲述客户端应用与服务器端功能需求和评分细则。

## 详细要求

### 1. 客户端



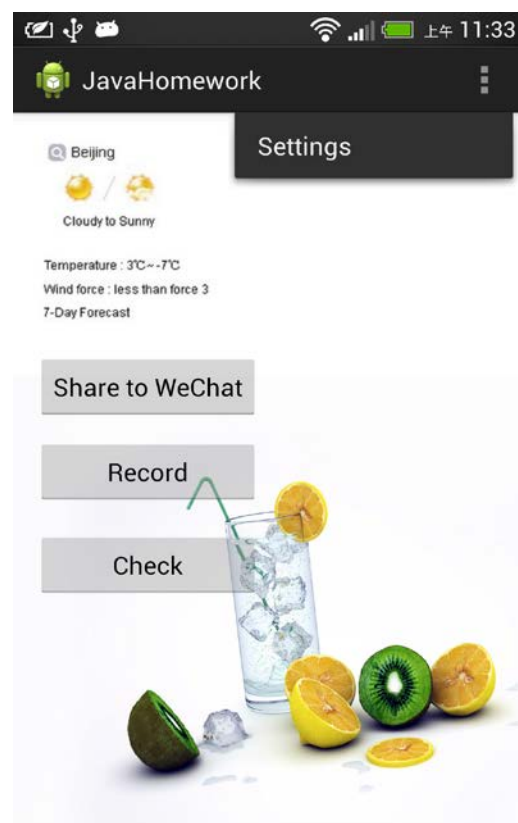
主要考察的知识点：

安卓手机客户端需要实现的功能：

- 采集手机数据（图像，GPS，wifi，时间，通话记录，时间）
- 将采集到的数据传输到服务器端存储
- 天气的显示
- 将信息分享给微信好友

下面将详细描述客户端应用各功能模块需求。

#### 1.1 首页天气显示



天气 API 可用附录一的 yahoo 天气，也可以使用其他 API。

注：

\*显示当日天气，温度，其他项可自行添加

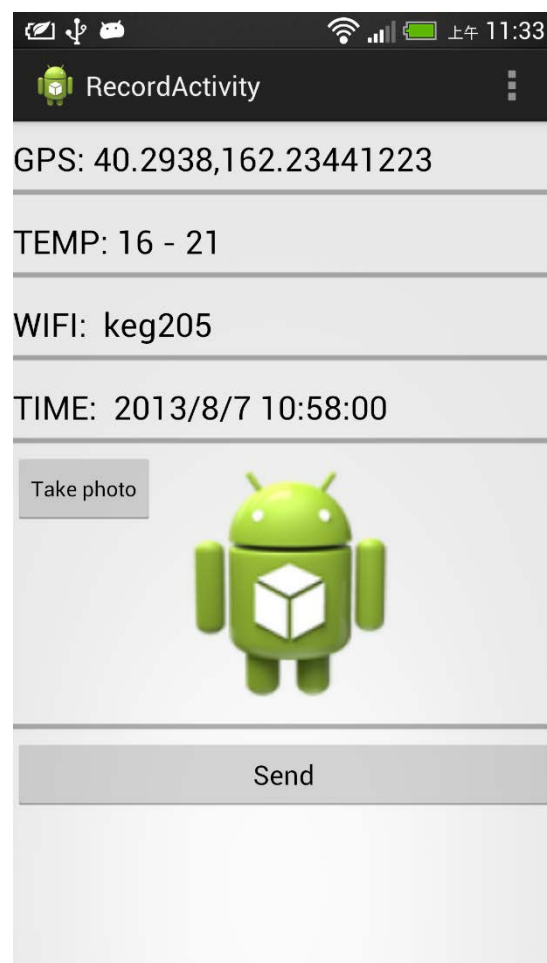
\*首页可进行适当美化，按照美观程度进行适当加分。

## 1.2 分享到微信

实现把天气信息分享到微信（发送给微信好友，分享到朋友圈）

注：点击分享按钮跳入新的界面，可对分享内容进行编辑，然后进行分享。

## 1.3 数据采集模块



点击 Record 进入数据采集界面，实现以下数据的采集：

- 图像：用手机拍摄当前照片
- GPS：当前位置的经纬度
- Wifi：测到的最强信号的 WIFI 信息
- 时间：采集时刻的时间戳
- 温度：当前气温

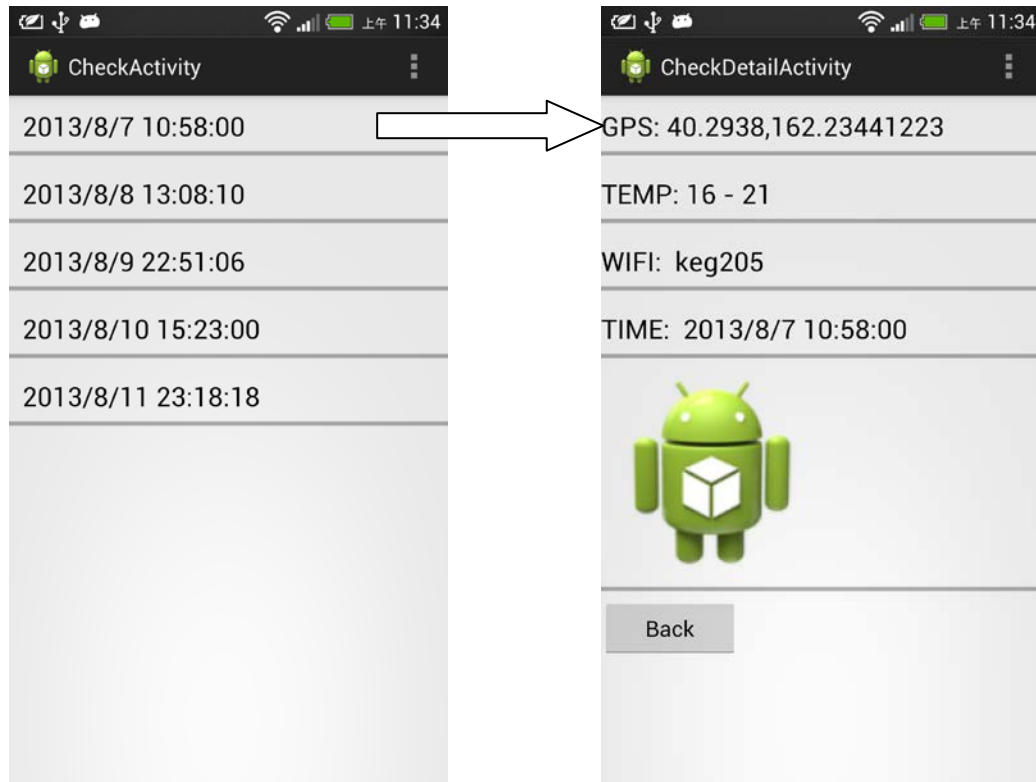
该界面显示最近一次采集到的数据。用户按下发送按钮可以将数据发送到服务器端。

注：

\*客户端每采集一组数据，随即把数据发送到服务器，采集不到的数据项以空或者用特定的符号（如“--”）代替。

\*照片拍摄后在图中 android 机器人位置显示

## 1.4 感知数据显示模块



从服务器端查询数据并在客户端显示，如上图所示。

左图为按照时间排列的记录条目，点击进入详细数据内容界面

注：

\*记录按照时间顺序显示，可以按照自己的想法扩展查询显示方式（若有实现则适当加分）。

## 1.5 设置

在客户端需要有配置页面，用户在上面可以配置的项有

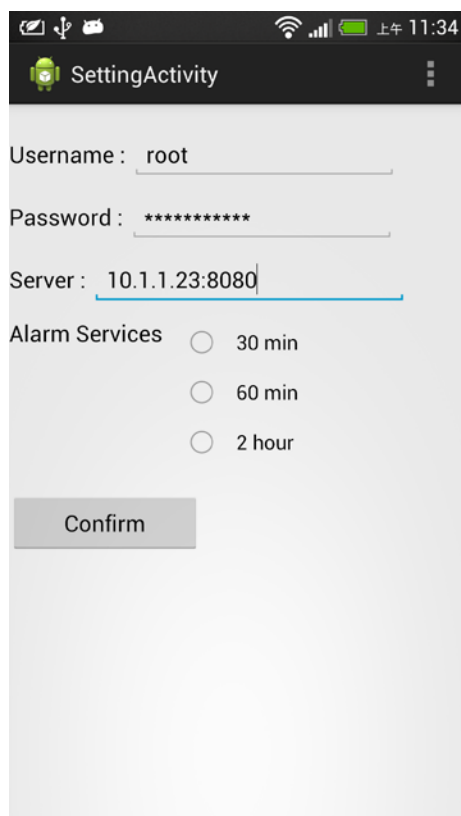
- 用户名
- 密码
- 服务器 IP 地址
- 闹铃提示，可选项有“每 30 分钟，每 1 小时，每 2 小时”，默认情况下“每 30 分钟”被选中。

注：

\*用户名，密码和服务端 IP 需要用来和服务端通信。

\*提醒闹铃为按照设定的时间发出提醒，然后进行数据采集并发送到服务器。

示例界面如下图所示。



The screenshot shows an Android application interface with a black title bar containing the text "SettingActivity" and an Android icon. The main content area is light gray and contains the following elements:

- A "Username" label followed by a text input field containing the text "root".
- A "Password" label followed by a text input field containing ten asterisks "\*\*\*\*\*".
- A "Server" label followed by a text input field containing the text "10.1.1.23:8080".
- A section titled "Alarm Services" followed by three radio button options: "30 min", "60 min", and "2 hour".
- A gray rectangular button labeled "Confirm" at the bottom.

The status bar at the top of the screen shows various icons including a USB connection, signal strength, battery level, and the time "上午 11:34".

## 2. 服务器端

服务器负责接收手机客户端的数据并存储，并开发一个 B/S 架构的网站来帮助用户管理自己的数据。针对服务器端的数据管理系统，要求实现如下几个功能模块：

### 2.1 与客户端通信模块

该模块负责与客户端通信，要求完成：

- 接收客户端发送 http 请求，服务器端的 servlet 解析该请求并取出请求中包含的数据，可以简单的使用字符串拼接来组装数据（具体形式由组内两名同学协商）。
- 服务器获取到的是客户端整理好的数据包，数据包的格式需要客户端和服务端协调设计，统一编解码方式，并且服务器需要合理设计数据包的存储方式，因为涉及到异构数据的整合，同学们可以开动脑筋，如何能够有效地管理异构数据。涉及到的用户行为数据有：

用户标识	● 用来标识该条数据产生的用户
时间戳	● 数据采集的时间戳
位置	● 经度&纬度 ● WiFi 强度 ● MAC 地址等
图片	● 用户自拍上传的 ● 可以是环境的，可以是自身的
语音	● 自己录制的一些内容
文字描述	● 用户自己对当前时刻的一个自主描述
环境数据	● 光强，温度，PM2.5，湿度等；

- 同时，该模块要负责根据客户端的请求对服务器的数据进行处理，并将处理结果返回到客户端（这部分要求参加客户端要求）

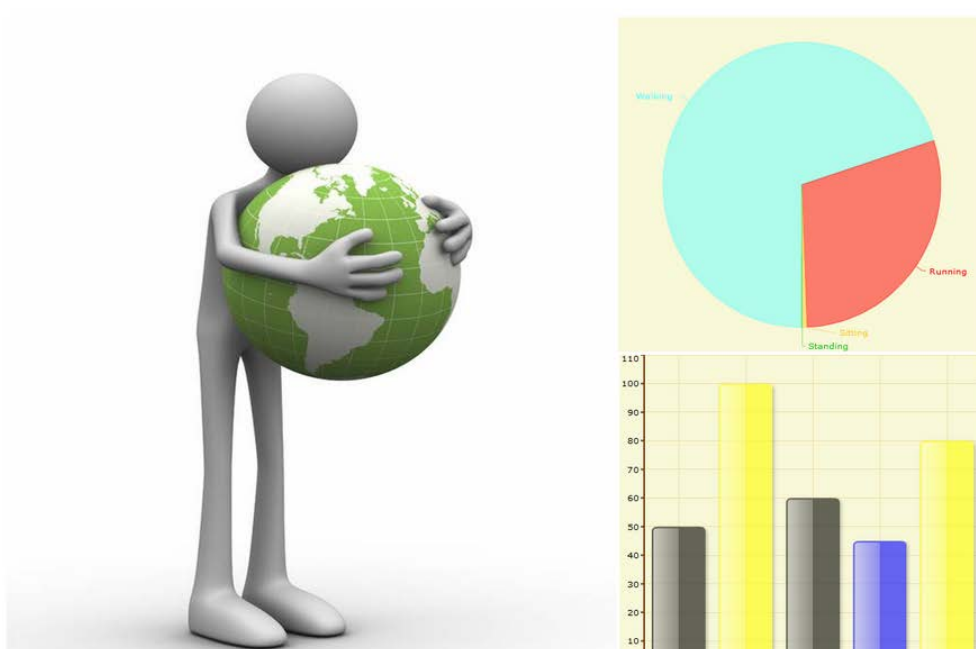
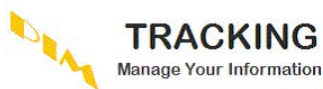
**说明：**因为每条数据是来自客户端的，所以该用户标识应该是客户端能够得到的。这部分不需要界面，用户标识可以用网站的用户名，对于网站而言，这个用户名是全局的；另外，可以在客户端获取用户手机的 deviceId（具体参照 Android 的 API）作为标识用户。其余的数据完全可以通过 API 获取到。

### 2.2 用户管理模块

本系统需要注册才能使用，开发网站的用户管理模块，包含的基本功能如下：

- 导航栏

下图是一示例，使用 HTML 来编写网页的导航栏，用来触发后续功能。通过点击导航栏的各项来跳转页面，进入各个功能子页面。



● 用户登录/登出

用户只有登录后才能查看自己的相关数据。设计登陆界面，点击“登录”进入登陆页面：

The login form is titled '会员登录' (Member Login) and includes a close button (X). It contains two input fields: '用户名' (Username) and '密码' (Password). Below the fields are two buttons: '登录' (Login) and '重置' (Reset).

要求：

- 包含如上图所示各元素，页面形式自定；
- 登录时检查用户名和密码是否为空（使用 javascript）；
- 登录失败提示重新登录，并返回登录页面；
- 登陆成功后直接跳转到“数据管理”页面；



- 页面欢迎信息设置，登陆成功后将用户名在网站上显示，并在用户退出前，能够随着页面跳转维持显示用户名，下例仅供参考，具体形式自行设计，不必完全拘泥于该例：



- 点击“退出”将注销 session，返回“首页”。

- 注册新用户

点击导航栏“注册”将跳转到注册页面：

## 注册

[访问](#) 清华大学知识工程实验室

要求：

- 注册对应的后台程序需要能够实现例如用户名查重功能，如果注册的用户名已存在，给出提示；
- 密码和确认密码的一致性验证；
- 注册成功后直接登录，与登陆成功后一样。

- 个人信息管理

点击导航栏的“个人信息管理”对于注册的新用户，提供一个页面为其填充详细的个人信息，例如：真实姓名，联系电话，家庭住址等。

要求：

- 用户登录后，点击“个人信息管理”后，该页面上直接显示个人信息

真实姓名	许斌
职称	副教授
所在研究所	软件所
办公室	东主楼10-204
学生	马远超, 白音, 崔健
E-mail	xubin@tsinghua.edu.cn

  
上传头像

修改

- 点击“**修改**”按钮后，跳转到修改信息页面，该页面形式与“**注册**”相似，字段不同而已，点击修改完后保存，然后再跳转回“**个人信息管理**”，显示修改的信息。
- 提供用户的头像上传功能，如上图所示，点击“上传头像按钮”

## 2.2 说明：

对于上述要求，根据课上学习的 **jsp** 相关知识（网上资料调研）编写前台实现相应功能。创建 **user.txt** 在服务器端用来保存用户信息，应该包含最基本的登录信息，例如用户名，密码等。另外创建 **info.txt** 来保存在“**个人信息管理**”填写的用户的详细个人信息，具体内容自己设定，对应于该页面上的内容即可。

前台可以使用 **jsp** 页面来与服务器端 **servlet** 进行交互通信，用户登录信息验证和注册新用户等功能通过对服务器端的数据文件的读写操作完成。

\*有能力的同学可以使用数据库来更好的完成数据的管理任务。同时，如果部分同学对网站开发有浓厚的兴趣，推荐使用一些开源框架来完成网站的构建工作，使用的同学将会获得额外加分。下面提供一些比较流行的开源框架：

前台： **jquery**

<http://jquery.com/>

业务逻辑控制： **struts2**

<http://struts.apache.org/2.2.1/index.html>

后台数据管理：

如果使用数据库的同学可以尝试使用 **Hibernate** 更好的操作数据库内容

<http://www.hibernate.org/>

基于上面的开源框架，有兴趣的同学可以尝试了解下 **spring** 框架相关内容：

<http://www.springsource.org/>

## 2.3 用户行为数据的展示与管理

客户端采集的是用户感知信息，然后将数据传输到服务器端进行存储。用户可以登录到网站展示和管理自己的数据。数据管理模块包含的功能如下：

- 数据展示功能：

客户端收集的用户数据包是以时间为序列的数据集合，如何合理的利用 Web 提供的丰富方式将用户的行为状态数据进行展示，需要同学们进行合理的设计，这里并不做硬性设计要求。但是需要将采集到的异构数据（数据类型见 2.1）都进行展示，展示方案自定。

但是需要满足如下要求：

- 用户收集的异构数据源必须围绕时间轴聚集，全面展示所要求的数据；  
即时间戳是每个用户的异构数据包的唯一标识；



- 网站上可以对任何资源增加描述信息，同时可以分享到新浪微博；
  - 因为是异构数据，同学可以分别分享资源；也可以将资源整合成一种你认为合适的形式，并分享，这里不做强制整合要求；
  - 分享方式可以参照：<http://open.weibo.com/wiki/ShareCode>
- 要求利用地图 API 展示自己的轨迹信息；
  - 可以利用 Google Map API 或者百度地图 API 开发；
- 可以单独浏览某一类资源，但是在主视角周围要提供其他同一时间的其他类型数据的入口；
  - 例如，同学可以浏览自己最近一段时间的上传的状态图片，每张图片下面需要用较小的空间放置于该图片一起上传的其他的的数据，如语音，文字描述，天气情况等信息；如果没有，预留出编辑框，可以添加。

- 数据管理：

“数据管理”页面示例如上图所示。用户成功登陆网站后自动跳转到该页面，

用户可以通过查询条件来浏览自己的数据。

**需要完成如下基本要求**（扩展可以加分）：

- 数据查询，可以根据时间段查询数据流：

### 查询

开始日期:  结束日期:    
数据类型:

时间	位置	Wi-Fi	加速度	光强	
3:44 PM/7/20/2012	39.9, 116.3	KEG205, 00:19:e0:c7:6d:50, -77	9.6, -2.2, 6.8	160.5	<input checked="" type="checkbox"/>
6:00 PM/7/20/2012	39.9, 116.3	Tsinghua-IPv6, 10:8c:cf:10:94:22, -91	-9.8, 1.1, 2.2	70.3	<input checked="" type="checkbox"/>
6:13 PM/7/20/2012	39.9, 116.3	KEG205, 00:19:e0:c7:6d:50, -80	-9.8, 3.1, 7.9	144.1	<input type="checkbox"/>

- 数据编辑：上面提到的可以对各种资源进行文字描述描述的添加、修改和删除；
- 数据删除：用户可以对各种异构数据进行删除操作；（有能力的同学可以尝试数据恢复，加分）

## 2.4 用户行为数据分析与挖掘

本次大作业特意设置了开放性科研题目，希望大家提早感受科研而不仅仅是编码的工作方式与带来的快乐。

需要完成的系统不仅仅要包括用户行为数据的采集和管理，还要进行初步的分析工作，因为这部门可能会涉及到一些领域的深度知识，所以不做结果要求，实现功能框架即可，有兴趣的同学可以深入做一点，检查的时候能说明道理就可加分。设置的要求如下：

- 用户行为关键字抽取：希望用户可以通过上传的文字描述等文本信息中抽取最关键的几个词，用来作为这一时间段的标注，显示在数据的周围。当然，也可以通过其他资源中抽取有趣的内容。
- LBS 相关分析：用户会采集地理位置信息（GPS），可以利用这些 GPS 上附带的速度数据和地理位置数据计算出用户一天的行程信息、轨迹信息等。同时，利用地图展示出用户最重要的地点（例如：最经常出现，停留时间最长等）。
- 异构数据源的关系：是否天气会影响你的心情？上传图片都是在室外活动的时候？语音信息大多数是发在晚上？这些规律你发现了么？

**提示：**上面这些深度分析的结果不要求有多准确，大家当做额外的兴趣即可。但是需要大家利用一些开源工具能够有效的展示出你分析出的结果，可以利用：

- 地图 API（上面提到的）来展示自己的分析结果，如热点信息；
- 利用可视化技术展示自己分析挖掘出的有趣统计信息，可以使用

Open Flash Chart : <http://teethgrinder.co.uk/open-flash-chart/>

Jfreechart : <http://www.jfree.org/jfreechart/>

来在自己的网站上进行折线图、饼图和柱状图的绘制。

## 2.5 Java 设计模式的应用

学习了 java，就应该在应用开发中了解并使用相应的设计模式，这样才会让自己的应用更好管理和扩展，本次大作业需要同学们使用 3 种 java 设计模式在自己开发的系统中，这里不做必须功能要求，完成了该部分要求的有额外加分。

可以参考：<http://www.ibm.com/developerworks/cn/java/design/>来了解各种设计模式的原理和应用。

## 完成方式与评分要求

- 分组要求
  - 本次大作业采用分组合作的方式完成，自由组合，2 人一组；
  - 1 人负责客户端，1 人负责服务器端，具体参见上面要求；
  - 最后分别为个人打分（客户端和服务端满分都是 100 分）
  - 最后成绩是给每个人一个最终成绩，考虑到服务器端和客户端合作的需求，最终成绩分为个人成绩（75%）和小组合作成绩（25%）；小组合作成绩组内两人是一致的，即占最后个人总分的 25%，后面的评分细则中将会指出哪部分是合作部分，小组合作成绩将通过 2 人各自的合作部分的平均分算出；
- 以下是详细评分准则：
  - 客户端

模块	子功能	百分比 (%)
与服务器端通信模块	发送数据到服务器	10
	发出请求并获取服务器端发来的数据	10
	无网络环境下的数据备份	5
天气模块	从天气源获取天气数据	10
	良好的界面显示	5
数据采集模块	采集数据的显示	5
查询显示模块	默认查询结果展示	10
	扩展查询方式	最多+10
分享到微信模块	可对发送内容进行编辑	5
	发送给微信好友	10
	分享到朋友圈	10
提醒功能	间隔闹钟提醒	10
其他扩展		最多+10
配置模块	设置模块	10
总计		100

满分 100 分，另外如果有适当的扩展则最多可以另加 20 分，即学生最多可以拿到 120 分。

➤ 服务器端：

服务器端各功能所占比例如下表所示，其中加底纹的部分是服务器端的合作部分：

模块	子功能	百分比 (%)
与客户端通信模块	接收并解析客户端请求，取出请求中包含的数据；	5
	处理并合理保存；	10
	根据客户端请求处理数据并返回结果	5
用户管理模块	导航栏	5
	用户登录/登出	10
	注册新用户	10
	个人信息管理	10
用户行为数据的展示与管理	数据展示	20
	数据展示	20
用户行为数据分析与挖掘	完成功能框架	5
总计		100

➤ 针对每个子功能：

◆ 基本功能完成：80%

◆ 页面美观：20%

➤ 附加分细则：（附加分最高为 20 分）

◆ 使用数据库代替文件系统进行数据管理：+2~4 分

◆ 使用开源框架：每种+1~3 分

◆ 数据可视化：进行一项数据可视化+1~2 分

◆ 用户数据分析挖掘：每项统计结果根据复杂度+1~1.5 分（最多 10 分）

◆ 使用设计模式：每个+1 分（最多 5 分）

## 附录一：

## yahoo 天气 API

API: <http://xml.weather.yahoo.com/forecastrss?w=2163866&u=c>

w 对应各个地方在 yahoo 数据库中的 WOEID, 如北京的 WOEID 是 2151330; 武汉为 2163866;

u 为温度的单位, c 表示摄氏度; f 表示华氏度

说明:

<pre> &lt;yweather:location city="Wuhan" region="" country="CH" /&gt; &lt;yweather:units temperature="C" distance="km" pressure="mb" speed="km/h" /&gt; &lt;yweather:wind chill="9" direction="0" speed="3.22" /&gt; &lt;yweather:atmosphere humidity="100" visibility="1.59" pressure="1015.92" rising="0" /&gt; &lt;yweather:astronomy sunrise="6:49 am" sunset="5:25 pm" /&gt; </pre>	<p>位置: 城市, 国家 单位: 温度单位 (摄氏), 间距单位, 气压单位, 风速单位 风: 冷风级别 9, 风向, 风速 3.22 大气情况: 湿度, 能见度, 压强, 天文: 日出 6:49am, 日落 5:25pm</p>
<pre> &lt;title&gt;Conditions for Wuhan, CH at 10:00 am CST&lt;/title&gt; &lt;geo:lat&gt;30.58&lt;/geo:lat&gt; &lt;geo:long&gt;114.27&lt;/geo:long&gt; &lt;link&gt;&lt;/link&gt; &lt;pubDate&gt;Tue, 16 Nov 2010 10:00 am CST&lt;/pubDate&gt; &lt;yweather:condition text="Fog" code="20" temp="9" date="Tue, 16 Nov 2010 10:00 am CST" /&gt; </pre>	<p>纬度 经度 链接 更新时间/查询时间 天气情况: text 天气情况 (雾), code 天气图标, temp 温度 9 摄氏度, date 日期</p>
<pre> &lt;yweather:forecast day="Tue" date="16 Nov 2010" low="6" high="16" text="Mostly Sunny" code="34" /&gt; &lt;yweather:forecast day="Wed" date="17 Nov 2010" low="8" high="19" text="Sunny" code="32" /&gt; </pre>	<p>天气预报: day 星期, date 日期, low 低温, high 高温, text 天气情况, code 天气图标</p>