# 研究生调研系统策划书

**清华大学研究生团委青年发展研究中心**

**二零一六年十月六日**

## 背景

科学化、民主化、制度化是现代组织决策的必然要求，也是党和国家不断强调的决策方式。做到决策民主化最重要的前提条件就是将决策建立在有效调研的成果之上。

目前清华大学研究生群体接近3万人，随着学生工作科学化的发展，新形势下研究生群体呈现出的不断变化的需求和提点，对研究生群体的调研需求不断增加。但目前的调研方式和渠道目前还无法满足日益增长的调研需求。调研工作进一步的发展的增长点在于调研的专业化、制度化、信息化。从专业化水平来说，许多重要的调研抽样存在较大的问题，比如朋友圈传播、熟人扩散。这样的抽样方式使得调研缺乏专业化的基础。从制度化水平来说，目前校内只有“基础信息调研”一项有组织、有支持的常规调研，而其他的调研缺乏有效的动员方式和支持保障。缺乏制度和激励将进一步影响调研的科学性。从信息化来说，目前的调研基本上都是通过网络问卷统计，但是统计的数据分散在各个部门，不利于重复利用，降低了多次调研的性价比。

在此背景之下，建立一个校内调研系统势在必行。此调研系统要建立一套专业的调研机制，保证全面、有效的抽样，设计用户愿意提供真实数据的激励机制，建立满足信息发放、收集、汇总的信息化系统。

为了更为高效、准确、安全、持续地获取研究生发展的情况，我们希望通过创建“研究生调研系统”，为今后的重要调研服务，实现调研专业化、制度化、信息化。

## 研究生调研系统

### 需求说明

本项目开展前期我们将根据性别、硕博、年级、政治面貌、院系等维度对研究生群体进行分层划分，再对各层进行随机抽样，共2000名，将这2000名研究生确定为调研样本，用来代表全校研究生全体。随后通过邮件征询其加入调研库、提供信息的意愿，确定入库样本。再进行抽样、确定意愿，循环往复直到达到2000名，以确保样本的代表性和信息提供意愿。随后通过调研平台完成调研的发起和信息收集，完成用户自发查询和激励发放。通过数据库连通样本库和调研平台，完成前后端的协同，完成调研系统的全局实现。

#### （1）研究生调研系统的更新

每学年开学前，根据当前入学和毕业情况对调研样本对象进行更新，确保研究生调研系统样本代表性，可以依据学校当前每年4次毕业的时间点进行定期维护。

#### （2）研究生调研系统的追踪

长期追踪样本的功能，研究生调研系统建立起来后，可以对样本进行长期追踪，包括毕业后研究生调研系统个体的就业与成长等各项信息。

### 实现路径

本项目所设计的研究生调研系统，其主要实现路径主要包括以下四个部分：样本库的建立与维护、数据库的设计与搭建、调研平台的设计与制作、网页后期美工及服务器后期维护。这些实现路径的具体内容将在后面三节中具体介绍。

## 样本库的建立与维护

### 样本库的建立模式

样本库的建立以随机采样为基础，根据不同院系、性别、年龄等基本信息计算出每一类群体应采样的人数。根据计算出每个院系每类群体的具体人数，通知院系负责推荐参与样本库的同学，对于有推荐的不符合要求的，或者推荐同学不愿意配合的，要求院系重新推荐。为了防止有些少数群体在随机采样时被忽视，应预留一些空间给该类群体，包括：

1）人数特别少的院系；

2）男女比例失衡以至于有一个性别并没有能采到样本的情况；

3）全校范围来说年龄特别大或者特别小的学生。

### 样本库的数据收集

在样本库的建立模式中已经提到，样本库是通过随机采样与少数群体手工采样的方法建立起来的。在建立起样本库之后，应进行样本库的数据收集。该项工作应下放到各个院系去做，让院系根据采样名单相应特征收集数据。在收到反馈信息之后，根据其信息质量，对于质量比较差的样本，应考虑重新采集数据或者重新推荐。值得指出的是，对于手动选择的样本，如果反馈质量不佳，应想办法做思想工作，因为这部分样本是无法替代的。

### 样本库的建立过程

#### （1）研究生群体的特征分析

通过查阅《清华大学统计资料简编（2014年度）》，其中各单位学生人数如下表所示，包含北京协和医学院－清华大学医学部的相关内容。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 院 系 | 合计 | 本科生 | 硕士生 | 博士生 |
| 建筑学院 | 1225 | 634 | 316 | 275 |
| 土木水利学院 | 1617 | 901 | 290 | 426 |
| 环境学院 | 1039 | 355 | 326 | 358 |
| 机械工程学院 | 4051 | 1908 | 1070 | 1073 |
| 其中：机械工程系 | 1200 | 610 | 227 | 363 |
| 精密仪器系 | 684 | 285 | 174 | 225 |
| 热能工程系 | 693 | 361 | 111 | 221 |
| 汽车工程系 | 705 | 376 | 155 | 174 |
| 工业工程系 | 769 | 276 | 403 | 90 |
| 航天航空学院 | 1097 | 514 | 160 | 423 |
| 信息科学技术学院 | 5644 | 2475 | 1719 | 1450 |
| 其中：电子工程系 | 1793 | 989 | 369 | 435 |
| 计算机科学与技术系 | 1522 | 549 | 448 | 525 |
| 自动化系 | 1211 | 605 | 292 | 314 |
| 微电子与纳电子学系 | 522 | 85 | 312 | 125 |
| 软件学院 | 596 | 247 | 298 | 51 |
| 理学院 | 2237 | 1092 | 307 | 838 |
| 其中：数学科学系 | 592 | 398 | 86 | 108 |
| 物理系 | 779 | 412 | 44 | 323 |
| 化学系 | 749 | 282 | 120 | 347 |
| 地球系统科学中心 | 117 | —— | 57 | 60 |
| 材料学院 | 1094 | 423 | 312 | 359 |
| 电机工程与应用电子技术系 | 977 | 529 | 181 | 267 |
| 工程物理系 | 1058 | 613 | 131 | 314 |
| 化学工程系 | 777 | 448 | 128 | 201 |
| 经济管理学院 | 3310 | 1065 | 1876 | 369 |
| 公共管理学院 | 1012 | —— | 864 | 148 |
| 马克思主义学院 | 118 | —— | 37 | 81 |
| 人文学院 | 1546 | 1119 | 189 | 238 |
| 社会科学学院 | 984 | 369 | 330 | 285 |
| 法学院 | 1547 | 430 | 932 | 185 |
| 新闻与传播学院 | 640 | 311 | 259 | 70 |
| 五道口金融学院 | 332 | —— | 246 | 86 |
| 美术学院 | 1780 | 1096 | 530 | 154 |
| 生命科学学院 | 1351 | 428 | 108 | 815 |

#### （2）研究生群体的按特征抽样及院系样本匹配

根据《清华大学统计资料简编（2014年度）》，以样本容量为2000人计，各院系的硕士和博士采样人数如下表所示。

|  |  |  |
| --- | --- | --- |
| **院系** | **硕士** | **博士** |
| 建筑学院 | 26 | 23 |
| 土木水利 | 24 | 36 |
| 环境 | 27 | 30 |
| 机械 | 19 | 30 |
| 精仪 | 15 | 19 |
| 热能 | 9 | 18 |
| 汽车 | 13 | 15 |
| 工业工程 | 34 | 8 |
| 航院 | 13 | 35 |
| 电子 | 31 | 36 |
| 计算机 | 37 | 44 |
| 自动化 | 24 | 26 |
| 微电 | 26 | 10 |
| 软件 | 25 | 4 |
| 数学 | 7 | 9 |
| 物理 | 4 | 27 |
| 化学 | 10 | 29 |
| 地学 | 5 | 5 |
| 材料 | 26 | 30 |
| 电机 | 15 | 22 |
| 工物 | 11 | 26 |
| 化工 | 11 | 17 |
| 经管 | 157 | 31 |
| 公管 | 72 | 12 |
| 马院 | 3 | 7 |
| 人文 | 16 | 20 |
| 社科 | 28 | 24 |
| 法学 | 78 | 15 |
| 新闻 | 22 | 6 |
| 五道口 | 21 | 7 |
| 美院 | 44 | 13 |
| 生科 | 9 | 68 |
| 医学 | 17 | 33 |
| 核研院 | 11 | 18 |
| 高研 | 0 | 4 |
| 教研 | 6 | 11 |
| 叉院 | 2 | 9 |
| 协和 | 162 | 161 |

（待讨论：苏世民书院这样的短期学生是否纳入调研数据库？）

根据上面的表格，督促学校的院系推荐调研样本。各院系在推荐时应考虑到本院系的男女比例和特殊人群。采样之后的数据收集方式参考样本库数据收集部分中的所述。

#### （3）研究生匹配样本的调研意愿分析与样本筛选

在样本库建立的过程中，设置调研意愿的选项作为参考，样本库的建立将优先考虑选取调研意愿强的研究生作为研究生调研系统的成员。事实上，样本库的样本筛选并不完全以此为标准，而是以调研问卷的实际完成情况为主要标准。因为有些人虽然并不愿意参与调研，但是还是会认真填写问卷，这对于一个干净的数据库是有利的。在调研之前，应强调本次调研的重要性和实际意义，以获得更好的调研效果。调研之后的样本筛选参考样本库数据收集部分中的陈述。

#### （4）研究生调研系统的相关性方差检验

在采样初期，我们可以将非特殊人群样本随机分为两个部分，并共享特殊人群样本。这样，相当于建立起了两个内部样本库，这两者在随机采样部分不同而共享特殊群体。在后期的调研过程中，在拿到调研数据之后，可以计算这两个样本库的相关性，如果相关性比较高的话，有理由相信，这样的采样是有代表性的。

### 样本库的维护过程

#### （1）制定研究生研究生调研系统成员的电子认证

为了提高参与调研的同学的积极性，进一步地，提高调研问卷的质量，对样本库的所有成员进行电子认证，如果预算允许，提供纪念品。

#### （2）研究生调研系统的反馈与更新等

每一次新生入学之后，计算从上次入学到本次入学期间，毕业的学生数量。我们需要从新入学的学生中采集这么多数量的新的样本，采集的方法和之前是一致的。如果新生中有以前已经毕业过的学生且之前就在样本库中，需要留心这样的情况，这样的样本可能是具有更大的意义的。

## 数据库的设计与搭建

### 信息库与问卷库

本着遵循数据库设计范式的原则，创建学生信息表，构建学生基本信息数据库。基本信息库可以和info后台数据库对接，根据选取的2000余名研究生调研样本的学号信息，分别将与该学生有关的所有的基本信息都导入基本信息数据库中，包括姓名、性别、在读学位、政治面貌、年级、学院、民族、籍贯、生日、学号、研究生类别等全面的信息。

用户在登陆研究生调研系统时应使用其在info上的账号和密码，因此后台数据库需要维护用户登录表，同样该表需要与info后台数据库相关的学生登陆表进行对接。

每次按照不同的主题向学生发放调研问卷后，根据问卷调研的反馈数据动态创建问卷表，构建调研问卷数据库，并与填写问卷同学的信息进行关联。数据库需要保留下每次开展的研究生调研问卷信息，在考虑到数据存储压力的同时，要保证研究生基本信息和调研信息的安全。

### 数据库选择

开发数据库应用，选择一个好的数据库是非常重要的。目前，商品化的数据库管理系统以关系型数据库为主导产品，技术比较成熟。面向对象的数据库管理系统虽然技术先进，数据库易于开发、维护，但尚未有成熟的产品。国际国内的主导关系型数据库管理系统有SQL Server、ORACLE、MYSQL、SYBASE、INFORMIX和DB2。

### 服务器选择与并发分析

#### （1）服务器选择

**阿里云：**阿里云数据库（Relational Database Service，简称RDS）基于飞天分布式系统和全SSD盘高性能存储，支持MySQL、SQL Server、PostgreSQL和PPAS（高度兼容Oracle）引擎，默认部署主备架构且提供了容灾、备份、恢复、监控、迁移等方面的全套解决案。对于没有专业DBA团队的企业来说，云数据库是个不错的选择。费用为MySQL源码维护版40.8元/月起。

**学校信息中心服务器：**清华大学计算机与信息管理中心的运行服务室的规划建设以及信息系统的运行维护与用户服务工作，建立了数字校园运维服务体系，管理着各类服务器、网络、安全等设备300多台，近100个信息系统，设备总容量近140TB的集中存储，为信息系统提供了“安全、稳定、高效”的运行环境，“集成、共享、准确”的数据环境，为4万多师生员工的学习、工作提供“及时、准确、贴切”的用户服务，使信息系统充分发挥对学校教学、科研、管理和服务的支撑作用，为清华大学建设世界一流大学做好信息化运行服务保障。

#### （2）服务器并发分析

校内学生填写问卷时段不同，因此不会引起高并发量，业务流相对平稳，大约每分钟50连接，数据库能够支持此并发量。

## 网页端的设计与制作

### 网页端设计目标

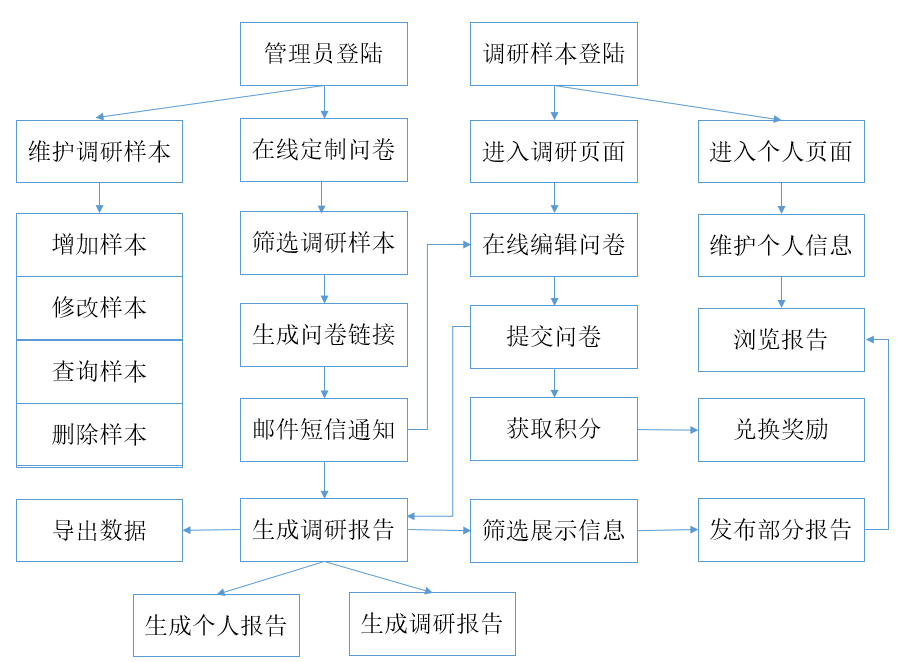
实现一个类似于问卷星的线上调研平台，不同之处在于需要定期维护全校2000余名学生的样本数据库（调研问卷的线上发放与填写主要是针对这2000余名同学展开，其他同学可选择性加入），并且可以考虑将调研结果经过统计分析以及交叉分析后以可视化的方式友好地展现给全校师生。与此同时，问卷的设计应该更有灵活性，在基础调研的基础上，可以针对不同特点的学生生成特色问卷，从而满足多样化的调研需求。

注：数据库的设计版本应该包含中英双语系统，如时间不允许，可以先开发中文系统投入使用，逐步加入英文系统，

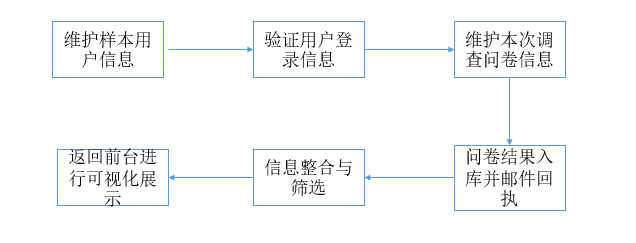
### 网页端工作流程

#### （1）基本业务流程

网页端前台基本流程如下。



网页端后台（依赖于数据库的搭建）基本流程。



#### 系统主要流程描述

##### A．管理员用户流程

管理员拥有系统的最高权限，其所具备的功能如下：

1. 维护样本数据库的样本信息，按照研究生的学号定期更新样本列表，保持调研的长期有效性。
2. 可以根据需要删除或添加样本属性字段，规定各属性字段的内容和格式。
3. 根据调研需求的不同，可以按照选取的条件筛选本次调研所面向的样本。可以根据每次调研需求的不同，对整体样本进行分组管理。
4. 查看单个样本每次问卷的完成情况，了解用户活跃度。
5. 在线设计调研问卷，生成调研问卷链接及二维码，通过短信和邮件实时通知样本用户完成调研工作并按时间回收问卷，将调研信息入库。
6. 按照既定的需要通过系统自动生成调研报告，管理员可以自行选择部分可公开的结果结论发布在前端，作为对调研样本的反馈。

##### B．样本用户流程

调研样本所具备的功能如下：

1. 通过info账号密码可以登录研究生调研系统，进入自己的主页，管理自己的相关基本信息。
2. 查看个人完成的问卷情况（参与时间、参与的调研、累积的积分等）。
3. 可以通过填写与自己相关的调研问卷获取相应的报酬和积分，通过积分可以领取奖品。（注：用户每次填写之后可以累计积分和经验值。经验值只增不减，用于辅助判断用户参与下一次调研可获得的收益。积分可以用于兑换金钱或者礼品）
4. 可以通过研究生调研系统浏览部分公开的研究生调研报告，了解相关调研的统计以及分析结果，获得信息的反馈。

### 网页端功能

#### （1）问卷编辑功能

该功能主要被系统管理员所使用。在发放问卷前系统管理员需要对问卷进行有针对性的设计，问卷可以由不同的题型组成，比如单项选择题，多项选择题，列表单选题，列表多选题，矩阵量表题，比重题，排序题，以及主观题等等。对于客观题来说，有制定好的模版，只需要系统管理员根据问卷选择拖动相应控件按钮并赋以文字含义即可；对于主观题来说，需要系统管理员根据需要手动输入设置，并限定样本用户输入规范等。

#### （2）问卷关闭功能

问卷发放前，管理员可以选择在某个时点或者填写人达到某个数量时关闭问卷，此时用户收到的链接失效，不可以再填写问卷获取相应积分。

管理员也可以在问卷发放过程中，根据实际情况，随时选择关闭问卷填写。

#### （3）数据分析功能

数据分析这一块主要由两方面内容组成，第一部分是将回收的样本用户的有效问卷内容转换成统计格式的数据，第二部分内容是将根据数据的含义分别采用不同的统计图进行可视化展示。在转换为统计格式的数据时，需要注意转化后数据的机器可读性，举个例子来说，为了方便机器进行进一步的可视化分析，对于多选题等结果数量不一定的题目可以采用一个01序列来表示，0表示未选中，1表示选中。问卷内容转化完成后，可以根据系统管理员的需要对不同类型的数据采用不同的统计图进行可视化展示，不同的统计图所突出的重点不一样，折线图突出一个变化趋势，柱状图突出数量的不同，饼状图突出不同种类数据的比例关系，概率分布图突出数据的分布情况，需要根据不同的需求选取不同的视图展示。

#### （4）邮件短信通知功能

邮件短信通知这块功能的设计相对来说比较简单。当系统管理员设计并生成好问卷后，可以点击前端页面按钮选项生成问卷链接以及一个独立的二维码，PC端用户和移动端用户可以通过点击链接或扫描二维码进入填写问卷。为了方便通知到用户，系统管理员可以按照条件筛选出符合条件的研究生样本，查询到他们的邮箱地址和手机号，并将该问卷的链接和二维码生成独立的邮件和短信群发给样本用户，让样本及时填写问卷。

#### （5）积分兑换功能

积分兑换功能主要为了提高样本用户和普通用户的参与热情而设计的。样本用户只要按时填写调查问卷，填写完成后就可以在网站领取相应积分，兑换相应的报酬或者礼品。与此同时，对于积极参与到活动中的普通用户也享有获取积分的权利，除此之外，普通用户可以向管理员提出相应的数据展示建议，如果得到采纳，也可以获得相应积分，同样的如果积分到达一定范围，可以兑换精美礼品。

## 研究生调研系统激励制度设计

对于参加问卷填写的同学，我们设置次结式和积分兑换制两种激励机制并综合使用，以提高其参与积极性，鼓励他们长期参与问卷填写工作。用户每次填写之后可以累计积分和经验值。经验值只增不减，用于辅助判断用户参与下一次调研可获得的收益。积分可以用于兑换金钱或者礼品

### 次结式

次结式即每次完成并提交问卷后立即可领取奖励，一般来说奖励以金额的方式发放。按照规定会定下一个“填写问卷最少金额”作为奖励发放的基准，根据问卷体量的不同以及调研样本活跃度的不同，完成问卷所获取的奖励金额也不同。对于问卷体量来说，基本原则是体量越大，相应获取的金额越多，对于相同体量的问卷一般来说奖励金额会根据当时的预算在平均值上下波动；对于调研样本活跃度来说，基本原则是用户历史参与过的调研次数越多，相应获取的金额越多，同样根据当时的预算在平均值上下波动。

次结算的结算方法：n表示用户填写问卷的次数，k表示问卷的题量（每5道题积1分），t表示在一定时间内完成答卷所获取的奖励积分（在12小时以内完成答卷，积10分；在24小时以内完成答卷，积5分；在1天以上完成答卷，积2分）。填写一次问卷可获得的次结金额为**account=（k+t）（1+5%）n**。

### 积分兑换式

积分兑换式即用户每次填写并提交问卷后，可以根据每次填写的题量获取相应的积分，当积分达到一定的数量时，可进行奖品兑换。积分兑换按照阶梯式划分等级。例如积分达到30分，可兑换30元手机充值卡，达到100分，可兑换U盘，达到1000分，可兑换kindle等，奖品的金额按照当时的预算而定。积分的增长速度取决于本次填写问卷的体量以及样本接受调研的频次，基本原则是问卷的体量越大，接受调研的频次越高，积分增长得越快。调研样本如果选择当积分达到某一等级时兑换奖品，则其积分相应扣掉该奖品对应的积分数。奖品的设置具有一定的跨越式，即越往后兑换奖品越大。

|  |  |
| --- | --- |
| **累计积分** | **奖品** |
| 20分 | 25元手机充值卡 |
| 30分 | 水杯 |
| 50分 | 32G U盘 |
| 70分 | 台灯 |
| 100分 | 64G U盘 |
| 150分 | 篮球/足球/羽毛球拍 |
| 200分 | 500G移动硬盘 |
| 500分 | Kindle |