

## ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ СОПРОВОЖДЕНИЯ ДЛЯ СОЦИАЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

*Романчуков С.В., Жаркова О.С., Берестнева Е.В., Маклакова Т.Г., Шаропин К.А.*  
(ФГАОУ ВО Национальный исследовательский Томский политехнический университет,  
Томск, Россия) e-mail: inoy@vtomske.ru

### INFORMATION SYSTEM CONCEPT FOR DATA COLLECTION PROCESS OPTIMISATION ON THE EARLY STAGES OF SOCIAL STUDIES

*Romanchukov S.V., Zharkova O.S., Berestneva E.V., Maklakova T.G., Sharopin K.A.*  
*National Research Tomsk Polytechnic University, Tomsk, Russia)*  
e-mail: inoy@vtomske.ru

In modern society it is difficult to overestimate the significance of sociological research, monitoring of the state of society and the dynamics of its change, especially if we take into account the current tense political and social context. However, such research raises a number of problems associated with obtaining the sample of respondents which has optimum size and displays the structure of the target population of people (residents of a particular region, employees of a region thereof, etc.) reliably. The research preparation process and data collection are sufficiently long and laborious, besides the action of interviewers on the ground and the collected material's quality should be quickly monitored to avoid significant selection bias or poor data quality and broken instructions. Such errors tend to accumulate, which confirms the need for their detection and correction in the shortest line. This work is devoted to the construction of a model of information system for social studies support at the early stages in order to ensure continuous quality monitoring and relevance control of the collected data.

Keywords: sampling, representativity, the structure of the population, statistics, modernization, interviewer quality control

**Введение.** Как и любой другой масштабный исследовательский проект, проведение социологическое исследование предполагает определенную последовательность этапов [1, 2].

На первом этапе формируется программа исследования, включающая цель исследования, его задачи, план, формулировку гипотезы на основе актуальности выбранной проблематики, определяются методы сбора информации, способы ее обработки, сроки проведения исследования и т.д.

На втором этапе социологического исследования осуществляется непосредственный сбор первичной информации. При этом могут применяться различные методы сбора информации — социологический опрос в форме анкетирования или интервьюирования; контент-анализ (записи исследователей, выписки из документов и другие сведения, полученные из различных источников документального характера); наблюдение, эксперимент и др.

На третьем этапе производится электронная обработка полученной информации, собранной в ходе социологического исследования на основе специальных компьютерных программ.

На четвертом, аналитическом, этапе проводятся анализ обработанной информации, подготовка научного отчета по итогам исследования, формулирование выводов и рекомендаций.[2]

При этом необходимо отметить, что зачастую самым трудоёмким и продолжительным процессом оказывается именно сбор первичных данных. Сам по себе процесс интервьюирования достаточного количества респондентов достаточно трудозатратен. Так, при подготовке

тысячи анкет, на каждую из которых у интервьюера уходит 30-50 минут (не считая времени, необходимого для прибытия к месту проведения опросов, выбора респондентов, получение разрешения на проведение беседы с конкретным респондентом и т.д.) только процесс интервьюирования требует 500-830 человеко-часов. Более того, на практике за счёт указанных причин это число можно смело удваивать. Представление результатов опросов в электронном виде и их подготовка к цифровой обработке также требуют определённых временных затрат. Кроме того собранные анкеты должны быть проверены на правильность заполнения и корректность представленных данных, что требует дополнительного времени, также необходимо проводить обзвоны интервьюеров, согласившихся предоставить свой номер телефона, с целью подтверждения самого факта проведения опросов. Однако, все проведённые работы могут оказаться недостаточными, т.к. после проверки отдельных анкет и их оцифровки может выясниться, что вся имеющаяся совокупность анкет, выборка, смещена относительно реального распределения, которое она должна отображать. В таком случае приходится либо проводить дополнительный сбор анкет с желаемыми параметрами, чтобы заменить им часть уже набранного материала, что вновь повышает трудоёмкость процесса, либо вводить весовые коэффициенты, что негативно сказывается на структуре выборки, изменяя её неслучайным образом (что для случайной выборки нежелательно) [3].

Всё это подводит нас к необходимости разработки и внедрения инструментария, позволяющего минимизировать трудозатраты на данном этапе, оптимизировать процесс проверки как отдельных анкет, так и самой выборки целиком, и динамически отслеживать изменения в выборке, в реальном времени корректируя задания респондентов при возникновении смещений. Данная работа посвящена разработке модели информационной системы, предназначенной для выполнения такого рода задач.

**Требования к системе.** Основной проблемой в проведении социологического исследования с использованием только традиционных методов является обилие излишней работы, которое ведёт к повышению трудозатрат и росту числа возможных ошибок в силу воздействия человеческого фактора. Интервьюер получает специально подготовленный бланк задания, выходит на маршрут, проводит интервью и заполняет бланк [3]. Этот документ проверяется человеком (экспертом), который принимает решение о допуске или недопуске полученной партии интервью в обработку. При этом на практике, как правило выбирается лишь несколько бланков из всех, предоставленных интервьюером (в целях экономии времени), что, однако, не гарантирует полностью отсутствия ошибок в остальных. После этого производится перевод содержимого анкеты в форму, пригодную для машинной обработки. После сбора достаточного количества анкет осуществляется проверка и (по необходимости) ремонт выборки.

Очевидно, что данную последовательность можно оптимизировать, используя возможности сети Интернет. Можно добиться существенного эффекта, развернув приложение, решающее ряд задач:

1. Динамическое обновление маршрутных листов и списков заданий интервьюеров (позволит сократить затраты времени на получение таковых в традиционной форме и повысить гибкость процесса);
2. Замена заполнения бумажных форм электронными;
3. Автоматизация проверки поступающих анкет (за счёт формальных критериев, заложенных в систему, спектр которых может быть обширен и превосходит возможности человека-проверяющего);
4. Использование данных системы позиционирования мобильного устройства, с которого осуществляется доступ к системе (с целью подтверждения факта присутствия интервьюера на месте проведения опроса)
5. Оперативное отслеживание состояния выборки по достижении некоторого существенного количества анкет и внесение изменений в задания и путевые листы интервьюеров;

Кроме того желательно, чтобы такого рода система содержала:

1. Формы обратной связи, в т.ч. в виде электронной конференции авторизованных участников;
2. Категориальный аппарат, методические указания;
3. Поддержка иных языков (как минимум - русского и английского);
4. Визуальное сопровождение, визуализация достигнутых результатов и заявленных целей (например с помощью интерактивных карт региона)[1].

С целью обеспечения безопасности и эффективной организации работы, сервис должен позволять разделение пользователей на различные категории, несущие различные функции и осуществлять разделение прав доступа различных категорий, гарантируя безопасность хранения информации и защиту от несанкционированного доступа. Для пользователей разных категорий целесообразно разработать отдельные интерфейсы, отражающие специфику их должностных обязанностей.

**Многоцелевой портал «MultiTest».** В рамках существующего на базе лаборатории Информационных технологий в социальных и медицинских исследованиях национального исследовательского Томского политехнического университета веб-портала «MultiTest», предназначенного для информационной поддержки психологических исследований, ведётся разработка такого рода системы как расширения уже существующего ресурса.

Для хранения информации портала применяется СУБД MySQL. Гибкость данной СУБД обеспечивается поддержкой большого количества типов таблиц и поддерживаемых форматов данных. В базе данных портала уже накоплен значительный объём информации о многолетних результатах исследований.

Непосредственно для описания тестов и методик используется универсальный формат представления данных XML, что позволяет, например, для создания новых тестов включаемых в портал, применять сторонние программы, в том числе обычные редакторы текста. В этом же формате осуществляется и хранение в базе данных информации о результатах тестирования[4].

Сама платформа представляет собой HTTP-сервер apache на платформе FreeBSD и реализуется с использованием php, html и AJAX. Портал доступен через сеть Internet, проведено его тестирование на наиболее распространённых браузерах (Internet Explorer v7 и выше, Mozilla v3 и выше Opera v9 и выше и т.д.).

Сервис позволяет разделение пользователей на различные категории, несущие различные функции и осуществлять разделение прав доступа различных категорий, гарантируя безопасность хранения информации и защиту от несанкционированного доступа.

**Заключение.** Разумеется, организация исследования с использованием подобных информационных ресурсов существенно отличается от классических «бумажных вариантов». Можно выделить ряд пунктов, по которым преимущества такой формы организации очевидны, но есть и такие, в которых превосходство новой системы неочевидно и требует дополнительных исследований для оценки.

В качестве преимуществ главным образом можно выделить следующие:

1. Анкеты собираются сразу в цифровой форме, нет затрат времени на их оцифровку (порядка 150-200 человеко-часов на 1000 анкет);
2. Снижаются временные затраты на проверку правильности заполнения анкет (конкретные объёмы требуют оценки);
3. Автоматическую проверку проходит каждая анкета, что эффективнее выборочной проверки нескольких экземпляров из партии;
4. Использование данных о расположении устройства, с которого была внесена запись, позволяет проверить факт присутствия интервьюера на заявленном месте с меньшими трудовыми затратами, чем традиционные обзвоны респондентов;

5. Использование возможностей сети интернет ускоряет выдачу заданий интервьюерам и их корректировку;
6. Сокращаются временные затраты на доставку анкет, получение бланков заданий и т.д.

С другой стороны функционирование подобной системы сопряжено с рядом неудобств:

1. Необходимость обеспечения постоянного функционирования сервера и оперативного устранения неполадок (что для традиционных форм менее критично);
2. Необходимы чёткие алгоритмы действия в случае нарушения соединения отдельного мобильного устройства, используемого интервьюером, с сетью (для сохранения результатов опроса в памяти устройства и их повторной отправки при восстановлении связи);
3. Проблема обеспечения безопасности и конфиденциальности данных (требует дополнительного изучения);
4. Вызванное ростом требуемых вычислительных мощностей удорожание оборудования и его обслуживания;
5. Стоимость использования мобильного интернета в отдалённых районах (требует дополнительной оценки);
6. Существенная стоимость проектирования и развёртывания такого рода систем.

Все эти эффекты нуждаются в дополнительном уточнении, однако предполагается, что положительный эффект от снижения трудозатрат на этапе сбора данных при проведении массивных социальных исследований и улучшении его организации превзойдёт возможные дополнительные затраты. Кроме того не исключено предоставление подобной системы, после её доработки и обкатки в нашей лаборатории, другим исследовательским группам.

***Исследование выполнено при финансовой поддержке РФФИ в рамках проекта № 14-06-00026.***

#### ЛИТЕРАТУРА

1. Zharkova O.S., Berestneva O.G., Moiseenko A.V., Marukhina O.V. Psychological computer testing based on multitest portal// World Applied Sciences Journal. - 2013. - Т. 24. - № 24. - С. 220-224.
2. Муратова Е.А., Берестнева О.Г. Выявление скрытых закономерностей в социально-психологических исследованиях// Известия Томского политехнического университета. - 2003. - Т. 306. - № 5. - С. 97-102.
3. Лапин Н.И., Беляева Л.А. Программа и типовой инструментарий «Социокультурный портрет региона России» (Модификация – 2010). - М.: ИФРАН, 2010. – 111с.
4. Моисеенко А. В., Берестнева О. Г., Щербаков Д.О. Развитие информационного ресурса для оценки компетентности ИТ-специалистов// Информационные и математические технологии в науке, технике, медицине: труды Всероссийской конференции с международным участием. – Томск. - 2012. – Т.2 - С. 12-14.
5. Марухина О.В., Берестнева О.Г. Анализ и обработка информации в задачах оценивания качества обучения студентов вуза// Известия Томского политехнического университета. 2004.- Т. 307. - № 4. - С. 136-141.