


МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ НАУК
Кафедра программного обеспечения

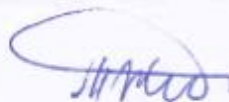
РЕКОМЕНДОВАНО К ЗАЩИТЕ В ГЭК
И ПРОВЕРЕНО НА ОБЪЕМ
ЗАИМСТВОВАНИЯ
Заведующий кафедрой
д.п.н., профессор
 И.Г. Захарова
29 июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРОВЕДЕНИЯ И АНАЛИЗА РЕЗУЛЬТАТОВ АНКЕТИРОВАНИЯ

02.04.03. Математическое обеспечение и администрирование информационных систем
Магистерская программа «Высокопроизводительные вычислительные системы»

Выполнила работу
Студентка 2 курса
очной формы обучения



Житенко
Татьяна
Сергеевна

Научный руководитель
д.п.н., профессор



Захарова
Ирина
Гелиевна

Рецензент
к.п.н., доцент



Плотоненко
Юрий
Анатольевич

Тюмень 2018

Аннотация

Магистерская диссертация – 62 с., 20 рис., 1 таблица, 3 главы, 29 источник, 4 приложения.

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРОВЕДЕНИЯ И АНАЛИЗА РЕЗУЛЬТАТОВ АНКЕТИРОВАНИЯ.

Объект исследования – возможности и средства автоматизированной системы. Цель работы – автоматизация процесса создания опросников, распознавания и статистических обработки результатов опроса, с помощью информационной системы поддержки проведения и анализа данных анкетирования. Результатом работы является разработанная система, включающая в себя подсистемы создания шаблона опросника, распознавания отмеченных анкет, и вывода статистики. Областью применения является исследовательские центры.

Оглавление

| | |
|---|----|
| Обозначения и сокращения | 3 |
| Введение | 5 |
| Глава 1. Автоматизация проведения социологических опросов и обработки анкетных данных | 7 |
| 1.1. Понятие социологического опроса как вида анализа социальной проблемы..... | 7 |
| 1.2. Современные средства автоматизации этапов проведения анкетирования..... | 10 |
| Глава 2. Обработка и анализ результатов анкетирования | 19 |
| 2.1. Методы и подходы обработки анкетных данных..... | 19 |
| 2.2. Статистическая обработка результатов анкетирования..... | 21 |
| 2.2.1. Описательная статистика..... | 23 |
| 2.2.2. Кластерный анализ..... | 24 |
| Глава 3. Проектирование и разработка автоматизированной системы поддержки проведения и анализа результатов анкетирования | 25 |
| 3.1. Архитектура системы поддержки проведения анкетирования..... | 25 |
| 3.2. Программная реализация автоматизированной системы поддержки проведения и анализа результатов анкетирования..... | 28 |
| 3.2.1. Создание шаблона опросника..... | 28 |
| 3.2.2. Распознавание анкет | 34 |
| 3.2.3. Анализ анкет | 39 |
| 3.2.3.1. Анализ результатов | 41 |
| 3.2.3.2. Кластерный анализ | 44 |
| Заключение | 47 |
| Список литературы | 48 |
| Приложение 1. Форма создания анкеты, директории работы с ней | 51 |
| Приложение 2. Форма записи вопроса и вариантов ответов | 52 |
| Приложение 3. Исходная анкета для кластеризации | 59 |
| Приложение 4. Словарь вариантов ответов для кластерного анализа | 61 |

Обозначения и сокращения

АС – Автоматизированная система.

ИМиКН – Институт математики и компьютерных наук.

ИПиП – Институт педагогики и психологии.

ФТИ – Физико-технический институт.

ОС – Операционная система.

РОС – Российское общество социологов.

СОМ – Компонентная объектная модель (Component Object Model).

OCR – Оптическое распознавание символов (Optical Character Recognition).

Введение

Исследования в области социологии являются неотъемлемой частью анализа социальной действительности. С помощью полученных результатов исследователи получают информацию самого разного рода, которая свидетельствует о процессах, явлениях, происходящих в жизни коллектива, в обществе в целом.

В проведении исследований в России и за рубежом получают широкое распространение онлайн-опросы [1]. К данной технологии прибегают как небольшие предприятия, изучающие удовлетворенность своих клиентов либо работу сотрудников, так и крупные исследовательские центры, опрашивающие население или большое количество людей. В сети распространено большое количество бесплатных и платных сервисов онлайн-опросов. Данные возможности таких сервисов экономят время, средства и человеческие ресурсы. И тем не менее, как и все социологические технологии, онлайн-опросы обладают определенными недостатками, которые делают невозможным их универсальное применение во всех сферах, где используется метод опроса [2]. В таких случаях применяются традиционные бумажные анкеты.

Преимущества бумажных анкет немало: факт личного контакта респондента с анкетой, мотивация заполнить все поля на бумажной анкете до конца, при таком анкетировании респондент может пропустить сложные вопросы, и вернуться к ним потом, или интервьюер может заметить низкую заполняемость анкеты и попросить ответить на пропущенные вопросы, что увеличит прозрачность исследования.

В свою очередь, традиционное («бумажное») анкетирование имеет свои минусы:

- Малый охват аудитории или значительные временные потери, особенно актуально для информации с других городов.
- Значительные трудовые и финансовые затраты (материалы для опроса, наблюдатели).

Также одной из проблем является анализ результатов опроса, так как большой объем данных занимает достаточно много времени для их оцифровки, и является слишком рутинной задачей для исследователей [3].

Поточное сканирование анкет – один из способов быстрой обработки результатов анкетирования широкого круга опрашиваемых. Сейчас поточные сканеры имеют в своем оснащении автоматическую подачу документов, что позволяет сократить ручной труд и достигнуть скорости оцифровки до 200 изображений в минуту на один сканер.

Существует достаточное количество программных продуктов для автоматизации процесса проведения отдельных этапов бумажного анкетирования, однако было бы целесообразно обеспечить комплексное решение наиболее важных задач: создание опросников, распознавание и анализ результатов.

Указанная проблема определила цель выпускной квалификационной работы – автоматизация процесса создания опросников, распознавания и статистических обработки результатов опроса, с помощью информационной системы поддержки проведения и анализа данных анкетирования.

Задачи:

1. Изучение особенностей проведения и обработки данных социологических опросов.
2. Изучение методов и алгоритмов распознавания специальных символов.
3. Проектирование и разработка системы для поддержки проведения и анализа данных результатов анкетирования.

Глава 1. Автоматизация проведения социологических опросов и обработки анкетных данных

1.1. Понятие социологического опроса как вида анализа социальной проблемы.

Социологический опросом называют метод получения первичной информации, основанием которого является непосредственная или опосредованная связь между опрашиваемым и респондентом с целью получения от последнего необходимых данных в форме ответов на вопросы.

На сегодняшний день, проведение опросов стало более широко распространенным, благодаря информационным технологиям в онлайн режиме – сервисами, позволяющими любому проводить исследования как бесплатно, так и с помощью специальных агентств. Но, несмотря на популярность онлайн опросов, многие исследователи считают полученные данные не всегда достоверными, по сравнению с традиционным бумажным анкетированием.

Не случайно предпосылкой для написания выпускной квалификационной работы на тему оказалось исследование, проводимое РОС, под научным руководством профессора Ю.Р. Вишневого, «Российское студенчество о Великой Отечественной войне» [4]. На сайте РОС были опубликованы как «бумажный» вариант анкеты (для скачивания и распечатки), так и ссылка на электронный вариант для онлайн-опроса [5].

Наиболее важные различия между бумажными и онлайн-анкетами заключались в следующем: первые оказались значительно более заполненными – интервьюеры убедились в том, что респонденты доходили до конца анкеты и то, что личный контакт с анкетой мотивировал респондентов на полное ее заполнение.

В анкетах онлайн около 20% респондентов не дошли до конца опроса, либо просто прекращали заполнять анкеты на середине, наиболее вероятно это было видно на открытом вопросе.

Студенты, ответившие на существенное количество вопросов в бумажной анкете, были заинтересованы в том, чтобы их труд не пропал, то есть старались дойти до конца анкеты. При «бумажном» анкетировании респондент может пропустить сложные вопросы в начале и вернуться к ним потом или интервьюер может заметить низкую заполняемость анкеты и попросить ответить на пропущенные вопросы. В ситуации с онлайн-опросом это невозможно, поскольку порядок вопросов важен для правильного определения переходов [6].

Именно поэтому опрос с бумажным анкетированием является наиболее часто используемым способом сбора информации в современных социологических исследованиях. Он представляет собой форму вербальной коммуникации между исследователем и респондентом [7].

Опрос может быть в свободной форме, когда опрашиваемые дают открытый ответ, - или же опросом, когда на вопрос можно ответить заранее определенным образом («да», «нет» или «незнаю»). Такой опрос с определенным методом ответа позволяет получать информацию в форме, удобной для статистической обработки, но также имеет недостаток, потому как сужает варианты ответов.

Одним из важных недостатков опроса является то, что с его помощью получают лишь субъективную картину, созданную опрашиваемым. Поэтому, чтобы минимизировать количество недостоверных ответов, исследователь должен уметь внушить опрашиваемому уверенность в том, что его ответы принесут пользу ему и окружающему миру. Ибо от опрашиваемого, от искренности его ответов во многом зависит качество опроса. А поэтому и вопросы должны быть хорошо продуманными и компетентными для респондента [8].

Главные части опросника – вступление, основная тема, переходы между темами, демографические вопросы и заключение [9].

Необходимо начинать опрос с короткого вступления, объясняющего респонденту, как, зачем, и для чего он был выбран для данного исследования:

- цель опроса;
- кто его проводит;
- добровольность участия респондентов;

- соблюдение конфиденциальности;
- время, необходимое для заполнения опроса.

Как правило, лучше всего начинать с общих вопросов и затем переходить к более конкретным вопросам [9].

- Хорошим тоном является приводить объективные факты, прежде чем задавать более субъективные вопросы.
- Следует убедиться, что ответ на предыдущий вопрос не влияет на то, как участник поймет следующий.

Использование переходов между группами вопросов, нужно для того, чтобы объяснить новую тему или формат. Например: «Следующие несколько вопросов касаются того, как часто вы смотрите телевизор. Пожалуйста, выберите ответ, лучше всего описывающий ситуацию».

К типичным демографическим вопросам относятся:

- пол;
- возраст;
- доход;
- национальность или место проживания;
- образование;
- расовая или этническая принадлежность [9].

В заключении следует поблагодарить участников за участие и объяснить ценность проводимого исследования.

Есть несколько хороших правил макетирования опросников:

- Формат опросника должен быть простым и понятным.
- Необходимо проверить, выполняют ли вопросы-фильтры свою функцию.
- Целесообразно делать опрос, не занимающий у респондента много времени.
- Оставлять достаточно места для ответов респондентов на открытые вопросы.

- Убедиться, что включены все подходящие варианты ответов для закрытых вопросов, в том числе «Другое» на случай, если респондент посчитает, что ни один из предложенных ответов ему не подходит. Там, где это уместно, включать ответы «Не применимо» и «Предпочитаю не отвечать».

- Убедиться, что вопросы плавно и логично переходят от одного к другому. Начинать опрос с общих вопросов, затем переходите к более конкретным или сложным моментам. Заканчивайте общими демографическими данными (возраст, пол и т.д.). Группировать вопросы по темам или разделам [10].

Хорошо, если можно показать прогресс респондентов по мере заполнения опросника. Это дает им ощущение продвижения и вдохновляет на продолжение процесса.

Следует протестировать опросник на нескольких респондентах до того, как он официально будет использован в исследовании. Такой тест позволит получить обратную связь и устранить возможные недочеты, касающиеся длительности опроса, технических проблем, формулировки и порядка вопросов.

1.2. Современные средства автоматизации этапов проведения анкетирования.

Крупное исследование, социологическое или маркетинговое, всегда требует немалых затрат. Для этого существуют специальные агентства, где уже обученные люди проводят опрос с помощью своего софта, который чаще всего не доступен широким массам, или используют планшеты, что тоже требует дополнительных вложений. Для небольших исследований достаточно программ и сервисов для проведения онлайн опросов, представленных в интернете. Такие сервисы позволяют провести онлайн анкетирование максимально просто: создается опрос, который распространяется среди пользователей, и на выходе получают обработанные результаты. Наиболее популярные из них: Testograf.ru, Survio.com, Surveymonkey.com, Google Forms, Simpoll.ru [11].

Testograf предлагает пользовательский функционал сервиса бесплатно, не

ограничивая при этом количество созданных опросов, вопросов и полученных ответов. Профессиональная версия, Testograf PRO, доступна платно. Есть возможность протестировать версию PRO, создав опрос без регистрации и оплаты, но, если есть необходимость сохранить опрос, нужно будет зарегистрироваться, а для получения доступа к результатам – оплатить.

Сервис оптимизирован для мобильных устройств. Кроме самостоятельного сбора ответов по прямой и дополнительным ссылкам, есть возможность встраивания опроса на сайт, всплывающее окно (автопоказ или клик по кнопке) и виджет опроса с настройками автопоказа и брендированием [11].

Предлагается также поиск респондентов по заданному таргетингу, при этом респонденты не являются оплаченными. Возможности настройки дизайна опроса довольно широки: цвет и шрифт текстов, цвет кнопок и чек-боксов, добавление логотипа и настройки шапки опроса, изменение фона (изображение, цвет). Сервер компании находится на территории РФ.

Survio.com. На бесплатной версии можно создать 5 анкет с неограниченным количеством вопросов и собрать макс. 100 ответов в месяц. Сервис оптимизирован для мобильных устройств. Помимо стандартного распространения опроса по прямой ссылке, есть встраивание опроса на сайт и всплывающее окно, однако обе этих возможности не имеют дополнительных настроек. Еще для сбора предлагается платное использование панели респондентов CINT. Настройки дизайна включают в себя: добавление логотипа, смену цветовой темы или использование готовых фонов. Сервер компании находится в Чехии [12].

Surveymonkey.com. Имеет пробную и бесплатную версии. Бесплатно можно создать 10 вопросов и собрать 100 ответов. Сервис оптимизирован для мобильных устройств. Распространять опрос для сбора ответов предлагается самостоятельно, возможностей несколько: прямая ссылка и дополнительные ссылки, встраивание опроса на сайт, всплывающее приглашение к опросу (есть настройка дизайна и % автопоказа), всплывающее окно самого опроса (с настройкой дизайна и % автопоказа). Сервер Surveymonkey располагается в США [12]. Дизайн опроса можно изменить с помощью подготовленных цветовых и фоновых тем либо использовать

пользовательскую тему с настройкой цвета и шрифта текстов. Разумеется, к опросу можно добавить логотип компании.

Google Forms. Многофункциональный сервис от компании Google, с сервером в США, позволяет бесплатно создать опрос, не ограничивая при этом создателя количеством опросов, вопросов и полученных ответов. Однако ответы придется собирать самостоятельно с помощью прямой ссылки или встраивания на сайт (без дополнительных настроек) [11]. Также имеется мобильная версия. При настройке дизайна опроса можно использовать подготовленные цветовые схемы, а также добавит шапку и логотип.

Simpoll.ru. Сервис имеет как пробную, так и бесплатную версии. Бесплатная версия позволяет создать 3 опроса, ограничивает число вопросов в одной анкете (макс.10) и число ответов на опрос (макс.100).Из минусов – отсутствие оптимизации для мобильных устройств. Респондентов также предлагается искать самостоятельно. Помимо прямой ссылки есть встраивание на сайт и всплывающее окно (без настроек) [11]. Настройки дизайна включают в себя изменение цвета/шрифта текстов и изменение фона опроса (цвет). Сервер сервиса находится в Эстонии.

Если рассматривать этап создание формы опросника, то наиболее популярное ПО для создания машиночитаемых анкет является ABBY FormDesigner.

Если сравнивать ABBYY FormDesigner с другими известными редакторами, которые можно использовать для создания макетов форм (MS Visio, CorelDraw, MS Word), то последние, предоставляя широкий спектр возможностей для создания разнообразных форм, тем не менее, специально не предназначены для создания именно машиночитаемых форм. Поэтому создателю формы приходится внимательно отслеживать соответствие формы требованиям машиночитаемости, требованиям к типу и внешнему виду формы, к типу и виду элементов формы, а также к расположению элементов формы. При соблюдении данных требований достигается максимальное качество распознавания текста, введенного в форму: вручную измерять размеры элементов, а также следить за их формой и расположением.

Программа ABBYY FormDesigner учитывает специфику разработки машиночитаемых форм, не требует дополнительного обучения операторов, имеет интуитивно-понятный пользовательский интерфейс и обладает широким спектром инструментов для создания и редактирования элементов машиночитаемой формы. Программа позволяет создавать как черно-белые, так и цветные фоновые формы. В программу встроена автоматическая проверка машиночитаемости. Проверка пригодности формы для автоматизированного ввода и дальнейшей обработки распознанного текста создаваемой формы, а элементы формы обладают набором свойств, созданных с учетом требований машиночитаемости [13].

Интерфейс программы ABBYY FormDesigner представлен на рисунке 1.

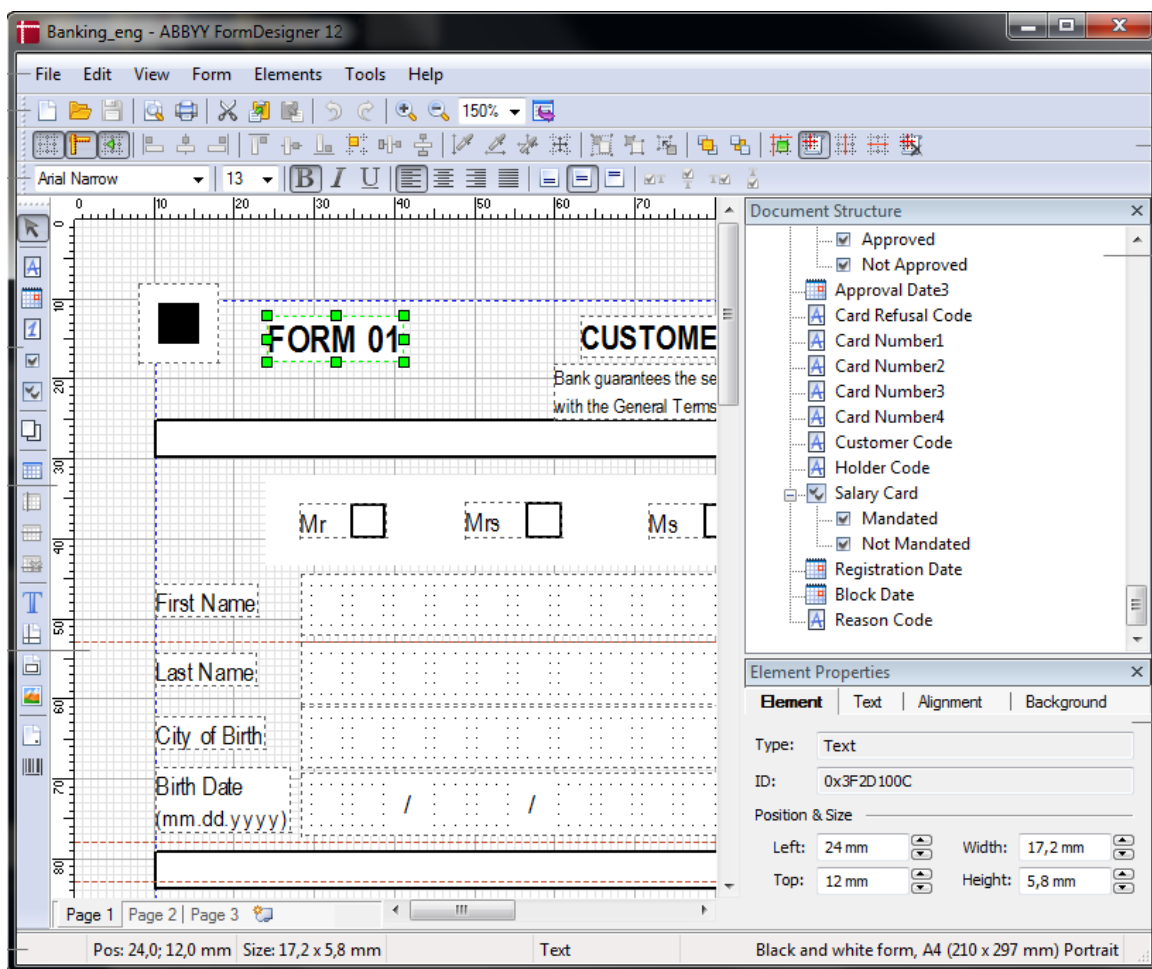


Рисунок 1.1 Интерфейс программы ABBYY FormDesigner

1.3. Машиночитаемые формы опросников

Изображение, представленное на рисунке 1.1 – это типичная заготовка машиночитаемой формы.

Анкеты, карточки социального страхования, опросные листы, бюллетени, гарантийные талоны и т.д. – все это формы, используемые для сбора разного рода информации [15].

Формы отличаются от документов следующим:

1. В отличие от обычных документов форма всегда имеет фиксированный набор полей.

2. В каждом поле формы может находиться информация строго определенного содержания: например, в поле "Фамилия" может находиться только фамилия (если, конечно, заполнивший форму не вписал туда по ошибке адрес), в поле Дата – дата и т.д.

Такой способ представления формы позволит получить информацию от большого числа респондентов с последующей компьютерной обработкой полученных данных. Ручной ввод информации в компьютер занимает достаточно много времени и человеческих ресурсов. Более того, такая форма работы ведет к большому количеству ошибок (опечаток), на исправление которых уходит много времени. Поэтому эффективнее автоматизировать процесс ввода форм [16].

Как правило, основной процесс автоматизированного ввода форм состоит из следующих этапов:

1. Настройка программы автоматизированного ввода на обработку формы данного образца (описание формы, выделение на образце формы данных для распознавания).

2. Создание изображения формы (сканирование).

3. Обработка изображений программой автоматизированного ввода (распознавание и проверка данных).

4. Экспорт данных во внешнюю информационную систему [16].

При обработке форм программа автоматизированного ввода сталкивается с рядом задач, от решения которых зависит качество распознавания данных из формы:

1. Определение положения элементов на изображении формы.
2. Отделение внесенной в форму информации (данных) от разметки текста, границ полей, фона, пояснительных надписей и т.д.

Чтобы облегчить программе решение этих задач, создаются специальные машиночитаемые формы. Для решения первой задачи машиночитаемым формам выдвигается требование точного соответствия формы образцу. Это значит, что у всех форм одного образца расположение элементов формы должно быть одинаковым (совпадать на просвет). Для решения второй задачи к машиночитаемым формам предъявляется требование специального дизайна формы: он должен позволять корректно отделить собственно данные от элементов формы.

Основными элементами, используемые в дипломной работе машиночитаемой формы будут текст и реперные точки [17].

Репер – это элемент машиночитаемой формы, необходимый для определения принадлежности формы к определенному образцу форм и ориентации формы во время автоматизированного ввода [18]. Первое, что определяется при распознавании данных из формы, это принадлежность формы к тому или другому образцу. Как уже говорилось, машиночитаемые формы создаются с учетом требования автоматизированного ввода, одним из которых является требование точного соответствия формы образцу. Количество и расположение реперов является основной характеристикой принадлежности формы к тому или другому образцу. Настроенная на работу с формами данного образца программа автоматизированного ввода ориентирует форму, т.е. пытается расположить её так, чтобы реперы совпали с реперами образца. Если реперы совпали, то программа может определить, к какому образцу относится данная форма и где находятся данные для распознавания.

Стандартные реперы приведены на рисунке 1.2.

| Черный квадрат | Уголок | Крест | Черный прямоугольник |
|---|---|---|---|
|  |  |  |  |

Рисунок 1.2. Реперные точки

Присутствие стандартных реперов на машиночитаемой форме крайне желательно, хотя в некоторых случаях в качестве репера может использоваться разделитель, описательный текст или штрих–код. Эти варианты реперов считаются приемлемыми, но по возможности их следует избегать, т.к. они могут привести к ухудшению распознавания формы.

Рекомендуется размещать на форме стандартные реперы (черные квадраты, уголки, кресты) в количестве 5 штук: четыре – по углам формы (при соединении центров этих реперов должен получаться прямоугольник) и один – на одной из сторон этого прямоугольника [18].

Как уже говорилось, машиночитаемые формы создаются с учетом требования автоматизированного ввода, одним из которых является требование к дизайну формы: дизайн формы должен позволять корректно отделить внесенную в форму информацию от разметки текста, границ полей, фона, пояснительных надписей и т.д. Существует два различных подхода для реализации этого требования:

1. Дизайн формы создается таким образом, чтобы при сканировании формы и последующей очистке изображения разметка текста, границы полей и фон исчезали.

2. Дизайн формы создается таким образом, чтобы на изображении формы после сканирования и очистки изображения были четко выделены разметка текста и границы полей [16].

При использовании первого подхода на изображении формы после сканирования и очистки остаются только реперы, описательный текст и информация, которую ввели при заполнении формы. Настроенная на работу с такими формами программа автоматизированного ввода будет искать на изображении данные, не ограниченные никакой разметкой, что сузит задачи программы и повысит качество распознавания. Для реализации этого подхода используются формы следующих видов: цветные фоновые формы, черно-белые формы с растровым фоном, черно-белые формы с растровыми границами, и черно-белые линейчатые формы. При использовании второго подхода на изображении формы после сканирования остаются реперы, описательный текст, информация, которую ввели при заполнении формы, а также разметка текста и границы полей [16].

Примеры машиночитаемых форм, представленные ниже, на рис.1.3 и рис.1.4, иллюстрируют использование специальных элементов, упрощающих распознавание вводимой информации.

The image shows a machine-readable form for a credit application from Sberbank. The form is green and white with a grid layout for data entry. It includes fields for personal data, contact information, and a date field.

СБЕРБАНК
Работа, деньги

Заявление-анкета на получение кредитного продукта

РОЛИ В ПРЕДСТАВЛЯЕМОЙ СДЕЛКЕ

Я ЗАКАЗЧИК И ОСНОВНОЙ ПОЛУЧАТЕЛЬ СРЕДСТВ
 Я СОЗДАМЩИК Я ПОРЯДОКЪЕЛ Я КЛИЕНТ Я ЗАЛОГОДАТЕЛЬ

ПЕРСОНАЛЬНЫЕ ДАННЫЕ

ФАМИЛИЯ: ИВАНОВА
 ИМЯ: ОЛЬГА
 ОТЧЕСТВО: ИВАНОВНА
 ДАТА РОЖДЕНИЯ: 02 10 1972
 МЕСТО РОЖДЕНИЯ: С. МОСКВА

ДАТА И МЕСТО РОЖДЕНИЯ

ДАТА РОЖДЕНИЯ: 05 10 2008
 ПЕРВАЯ БУКВА ФАМИЛИИ: ПЕТРОВА
 ПЕРВАЯ БУКВА ОТЧЕСТВА: []
 ПЕРВАЯ БУКВА ИМЕНИ: []
 ПУМКА ИНОЕ (укажите): []

КОНТАКТНАЯ ИНФОРМАЦИЯ

МОБИЛЬНЫЙ ТЕЛЕФОН (включая код оператора связи): +71 000 0000000

Рисунок 1.3. Машиночитаемая форма опросника для клиентов банка

В данной форме, помимо текста и реперных точек, использованы поля ввода, разделители, логотип предприятия.

Если при распознавании невозможно однозначно определить по реперам принадлежность формы к образцу (т.е. реперы у нескольких форм различных образцов совпадают), то для точного определения используется идентификатор в виде штрих-кода [13].

GR FORM Management Information System (MIS)
 Directorate General of Health Services
 Mohakhali, Dhaka, Tel: 8802 8816459 Fax: 8802 8813875, Email: info@dghs.gov.bd

1. Household Code: DIVISION, DISTRICT, UPAZILLA, UNION, WARD, HOUSEHOLD. Date: DD MM YYYY

2. Drinking Water Source: Tube well (Red) Tube well (Green) Tube well (Arsenic not tested) Rain Water River/Canal TAP Well/Draw-well Pond/Puddle Others

3. Latrine: Sanitary Pucca Raw Open field Bush-Jungle Others. Write in English

4. Economic situation: Unending hardship Temporary hardship Equal income-expenditure Solvent Can save. Approximate monthly spending

5. Family's main cell number: a) b) c)

6. Household Member 1: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

7. Household Member 2: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

8. Household Member 3: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

9. Household Member 4: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

10. Household Member 5: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

11. Household Member 6: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

12. Household Member 7: Serial, Date of birth, Sex, Marital status, Education, Occupation, Religion, Name, National ID, Chronic diseases (Poly-urea, High pressure, Cancer, Respiratory disease, Mental illness, Obesity), Date of death (From 01 January 2009), Cause of death.

Total Members: Filled by: More pages: Yes If yes specify page number

Рисунок 1.4. Машиночитаемая форма медицинской переписи

Глава 2. Обработка и анализ результатов анкетирования

2.1. Методы и подходы обработки анкетных данных.

Обработка и анализ информации, по мнению В.А. Ядова (выдающегося отечественного социолога, автора многих работ по методике и методологии исследований), – самый увлекательный этап исследования [19]. Обработывая, преобразуя полученные данные из разрозненного сырого материала, можно получить некоторую картину, выявляющую связи и закономерности в обществе.

Так, например, Мартышенко С.Н., Мартышенко Н.С., Кустов Д.А. в своей работе «Совершенствование математического и программного обеспечения обработки первичных данных в экономических и социологических исследованиях» рассматривают актуальность проведения социологических исследований не только среди научного общества, но и среди практикующих специалистов на предприятиях. В исследовании работа построена с соблюдением принципов системного подхода – совокупность подсистемы сбора первичных данных и анализа данных. Новый комплекс программных средств был создан на базе Excel, а также имеет интеграцию с другими пакетами статистического анализа (STATISTICA, SPSS, STATGRAPHICS и т.д.). Программный комплекс авторы рассматривают с точки зрения «проекта» и «модели данных» анкетного опроса. Все функции по сопровождению проекта на себя берет «мастер сопровождения», структура которого включает в себя: исходные данные, параметры проекта, даты изменений, фильтры, отчеты и изъятые данные. Модель данных определяет правила представления анкетных данных в проекте [20].

В.В. Арлазаров, В.Е. Кривцов, Д.В. Полевой, Д.Г. Слугин, И.М. Янишевский в своей работе «От графического образа к универсальному представлению формы документа», представили модель формы стандартных документов, объединяющий в себе бумажный и электронный документооборот. Они дают свое определение формы – это логический образ документа. В работе подробно описано следующее разделение формы на три модели: модель содержания, модель визуализации и модель взаимодействия с пользователем. Следующее разделение можно пояснить

так: один и тот же объект в разных процессах может выглядеть различно, но содержание должно быть одинаковым. Более подробно рассматривается модель взаимодействия на примере секционной модели формы. Модель представляет из себя ориентированный граф, где его вершины – это схемы секции, а ребра схемы переходников к следующей секции [21].

В работе Григорьева В.К., Грушина А.В «Особенности распознавания информации об успеваемости и анкетировании студентов» рассматриваются такие алгоритмы распознавания специфической информации, как «Алгоритм распознавания с позиционированием, основанным на опорной полосе» и «Алгоритм распознавания с позиционированием, основанным на использовании реперных меток».

В качестве объекта распознавания была выбрана зачетно-экзаменационная ведомость студентов, приведенная на рисунке 2.1.

The diagram shows a form for a Moscow State Institute of Radio Engineering, Electronics and Automation (Technical University). The form is titled 'ЗАЧЕТНО-ЭКЗАМЕНАЦИОННАЯ ВЕДОМОСТЬ' (Exam Record) for the subject 'Математический анализ (Экзамен)'. It includes fields for faculty (ВМС), group (ВБ-1-06), and semester (3). A table lists three students with their IDs, names, dates, and scores. The form also has fields for the examiner (Экзаменатор: RAL), the dean (Декан), and dates for the exam (Date of exam: KMM, Day of exam: KMD). A barcode (BC) and a support bar (OP) are also present.

| № | Ф.И.О. студента | Шифр | Дата | Оценка | для машинной обработки | | | | Подпись экзаменатора | |
|---|-----------------|--------|----------|---------|------------------------|-----|------|-----|----------------------|-----|
| | | | | | и.я. | зач | исуд | удо | | хор |
| 1 | Валиев Т.С. | 061001 | 19.10.07 | хорошо | | | | | | |
| 2 | Дормаков Е.Н. | 061003 | 19.10.07 | отлично | | | | | | |
| 3 | Доронкина А.В. | 061004 | 19.10.07 | отлично | | | | | | |

Рисунок 2.1. Расширенная модель зачетно - экзаменационной ведомости.

$Md=F(\text{Header},OP,BC,SL,RL,RAL,ES,EN,KMD,KMM)$, где BC – штрихкод, RAL – поле оценок для машинной обработки, KMD – день проведения контрольного мероприятия (КМ), поле для машинной обработки, KMM – месяц проведения КМ, поле для машинной обработки, OP – опорная полоса.

В случае алгоритма распознавания с помощью опорной полосы, в исследовании была использована надежная и очень широко распространенная в мире штрихкодированная технология.

Отсканированное изображение проходит проверку на минимально допустимое разрешение. В правом верхнем углу расположена опорная полоса, расположенная в правом верхнем углу документа, и по ее положению осуществляется сдвиг и поворот изображения. После считывания штрихкода, из базы данных загружается информация, и по ней производится генерация прототипа документа. Производится наложение прототипа на отсканированное изображение и оценка степени совпадения путем анализа интегрального приграничного шума.

Предложенные алгоритмы были экспериментально исследованы и подтвердили свою работоспособность. Данное программное решение зарегистрировано в отраслевом фонде алгоритмов и программ [22].

2.2. Статистическая обработка результатов анкетирования.

Одно из направлений использования методов математической статистики в социологии – разработка методов анализа данных. Данное направление стремится к поиску закономерностей в статистике, когда исходные данные не удовлетворяют строгим статистическим правилам. Их необходимость имеет место быть, если есть возможность формализовать характеристики изучаемых объектов, процессов и явлений [19].

В распоряжении исследователя есть некоторые измеренные числовые данные. Любое измерение представляет собой соотнесение значения измеряемого параметра с некоторой реальной величиной – эталоном. При этом эталон действительно должен быть именно эталоном (точная мера, служащая для воспроизведения, хранения и передачи единицы измерения какой-либо величины).

После соотношения измеряемых параметров и эталона, вычисляется оценка результата измерения. И если в естественных науках и технике существуют

материальные эталоны (длины, времени, массы, информации и др.), то в измерениях, выполняемых в социологических исследованиях, эталоны часто носят условный характер. Большинство характеристик изучаемых явлений и процессов не поддаются измерению, поскольку для них просто нет объективных эталонов, без которых нельзя выполнить корректное измерение [23]. Действительно, трудно представить эталоны мотивации, активности, пассивности, креативности и т.д.

Таким образом, в контексте выполнения измерений (для последующего привлечения математических методов) в исследовании можно выделить следующие этапы:

1. Выбор изучаемого явления (объекта, процесса), которому в процессе наблюдений будут даваться оценки в интересующем исследователя разрезе (в соответствии с используемой моделью).

2. Выделение (в виде критериев) существенных признаков явления, по которым его можно узнать, отличить от другого и представить, как объект измерения. Определение показателей выраженности каждого из выделенных признаков.

3. Выбор из числа существующих или разработка авторской методики для выполнения измерений, включая выбор шкалы измерений для каждого показателя [24].

Применение математических методов в исследованиях преимущественно связано с двумя направлениями: 1) обработка экспериментальных данных методами математической статистики; 2) построение математических моделей, описывающих изучаемый процесс, где, как правило, также используются подходы математической статистики, например, для прогнозирования на основе регрессионной модели.

Среди методов математической статистики различают параметрические и непараметрические методы. Если данные, подлежащие обработке, дают распределение частот близкое к нормальному, то возможно применение параметрических методов, базирующихся на свойствах распределения Гаусса. Методы, применяющиеся в случае несоблюдения нормального распределения или

малого числа наблюдений (это не позволяет оценить вид распределения), относят к непараметрическим.

Как и любые математические методы, методы математической статистики нельзя рассматривать в отрыве от конкретных типов задач, для решения которых они и предназначены. В числе основных прикладных задач математической статистики обычно выделяют следующие: 1) описательная статистика; 2) сравнение выборок и рядов показателей; 3) кластерный анализ.

2.2.1. Описательная статистика

Если первичный анализ исходных данных показывает, что распределение близко к нормальному, возможно применение параметрического метода. Для описания данных в этом случае вычисляют математическое ожидание (среднее значение) и дисперсию [23]. Дисперсия характеризует разброс случайной величины около среднего значения. При нормальном распределении случайной величины ее наблюдаемые значения с вероятностью 0,997 отклоняются от среднего значения не более чем на утроенную величину среднеквадратичного отклонения (правило трех сигм). Необходимо помнить, что оценка математического ожидания по выборке является случайной величиной, которая описывается распределением Стьюдента. Выборочная дисперсия также является случайной величиной (закон распределения хи-квадрат Пирсона). На основании распределения хи-квадрат строятся доверительные интервалы случайных величин. Если же характер исходных данных далек от нормального распределения, необходимо использовать непараметрические методы.

Примечание. Как правило, параметрический метод дает корректные результаты, если данные измерены с помощью шкалы отношений, т.е. существует неизменная единица измерений и точка отсчета. Однако, достаточно часто, когда это условие не выполняется, ошибочно используют для характеристики выборки среднее значение, что приводит к неуместным выводам. Наиболее характерным

примером служит сравнение качества обучения учащихся (или целых классов), отнесенных к контрольным и экспериментальным группам, на основании средних оценок. Это некорректный способ, поскольку отсутствует и абсолютная точка отсчета, и единица измерений, поэтому даже у одного педагога одна и та же оценка у разных учащихся свидетельствует только о том, что они принадлежат к определенной группе с точки зрения их обученности. В этом случае характеризовать выборку можно с помощью ранжирования, подсчета количества учащихся в выделенных группах и т.д.

2.2.2. Кластерный анализ

Кластерный анализ разбивает совокупность объектов на однородные классы (или кластеры). Это задача многомерной классификации данных [25]. Существует около 100 алгоритмов кластерного анализа, для дипломной работы выбран метод k-Means(к-средних).

Алгоритм k-средних принимает в качестве входных данных набор данных X , содержащий N точек, а также параметр K , задающий требуемое количество кластеров. На выходе получаем набор из K центроидов кластеров, кроме того, всем точкам множества X присваиваются метки, относящие их к определенному кластеру. Все точки в пределах данного кластера расположены ближе к своему центроиду, чем к любому другому центроиду.

Алгоритм делится на следующие этапы:

1. Исследователь выбирает число k -кластеров.
2. Программа соотносит каждый объект к кластеру с ближайшим «центром масс»;
3. Идет пересчет «центров масс» кластеров согласно их текущему составу.
4. Повторить п.2, если критерий остановки алгоритма не удовлетворен.

Зачастую результатом статистической обработки данных эксперимента становятся многочисленные графики и диаграммы различного типа. Достоверность эмпирического материала должна обеспечиваться выбором адекватных задачам исследования методов, подбором или разработкой корректных методик и шкал измерения, тщательностью проведения измерений в строгом соответствии с выбранной методикой измерения, точностью выполнения расчетов (как правило, с помощью специализированных программных продуктов).

Глава 3. Проектирование и разработка автоматизированной системы поддержки проведения и анализа результатов анкетирования

3.1. Архитектура системы поддержки проведения анкетирования.

Диаграмма системы выполнена с помощью языка моделирования бизнес-процессов Archimate. ArchiMate – язык архитектурного описания корпоративных и инженерных систем (моделирования архитектуры предприятия) [26].

Есть исследователь, который может играть роль как создателя анкеты, так и аналитика данных. Через интерфейс приложения он выполняет процесс проведения исследования, то есть работает с анкетой, собирает и распознает результаты и проводит аналитику данных.

Само приложение разбивается на три подсистемы: подсистема создания опросника, подсистема распознавания анкет, подсистема вывода результатов. Каждая подсистема выполняет свои функции создания, распознавания и аналитики анкет. Слой приложения взаимодействует с бизнес слоем.

Бизнес слой (слой предметной области) представлен на рисунке 3.1.

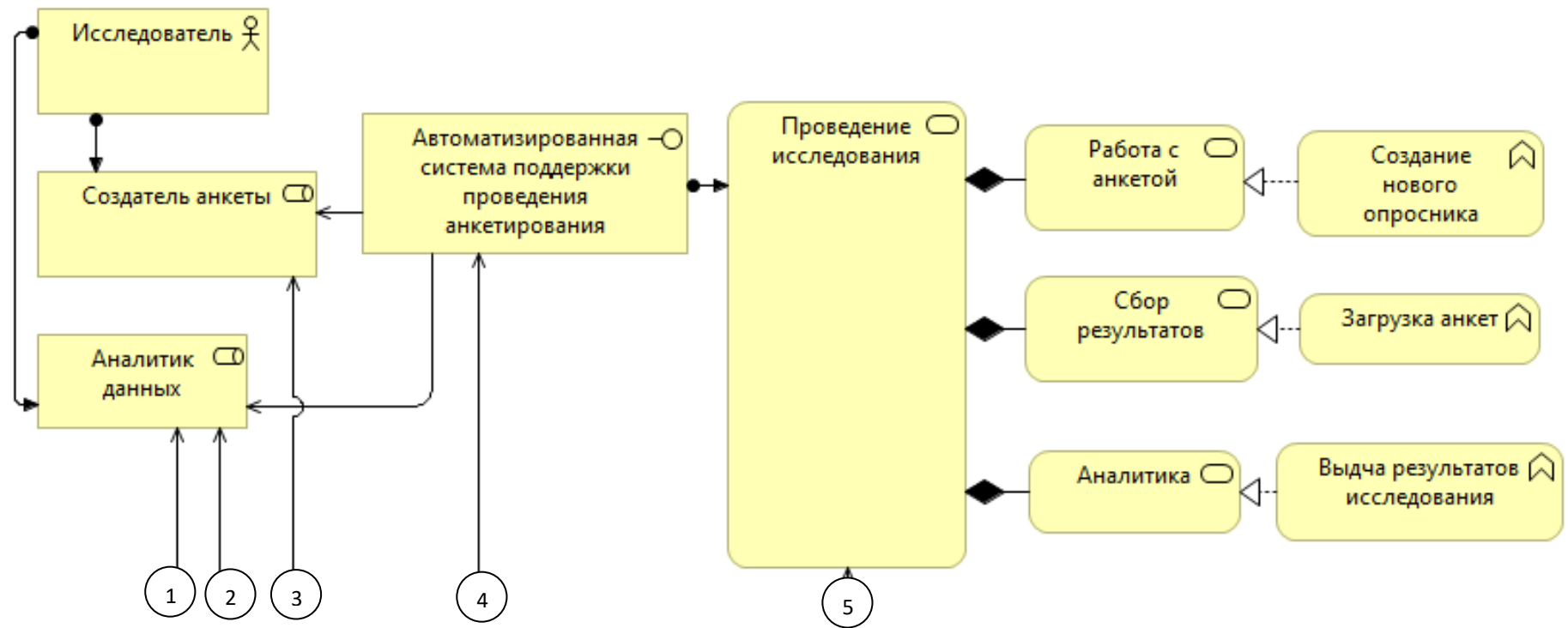


Рисунок 3.1 Бизнес слой системы

Слой приложения представлен на рисунке 3.2.

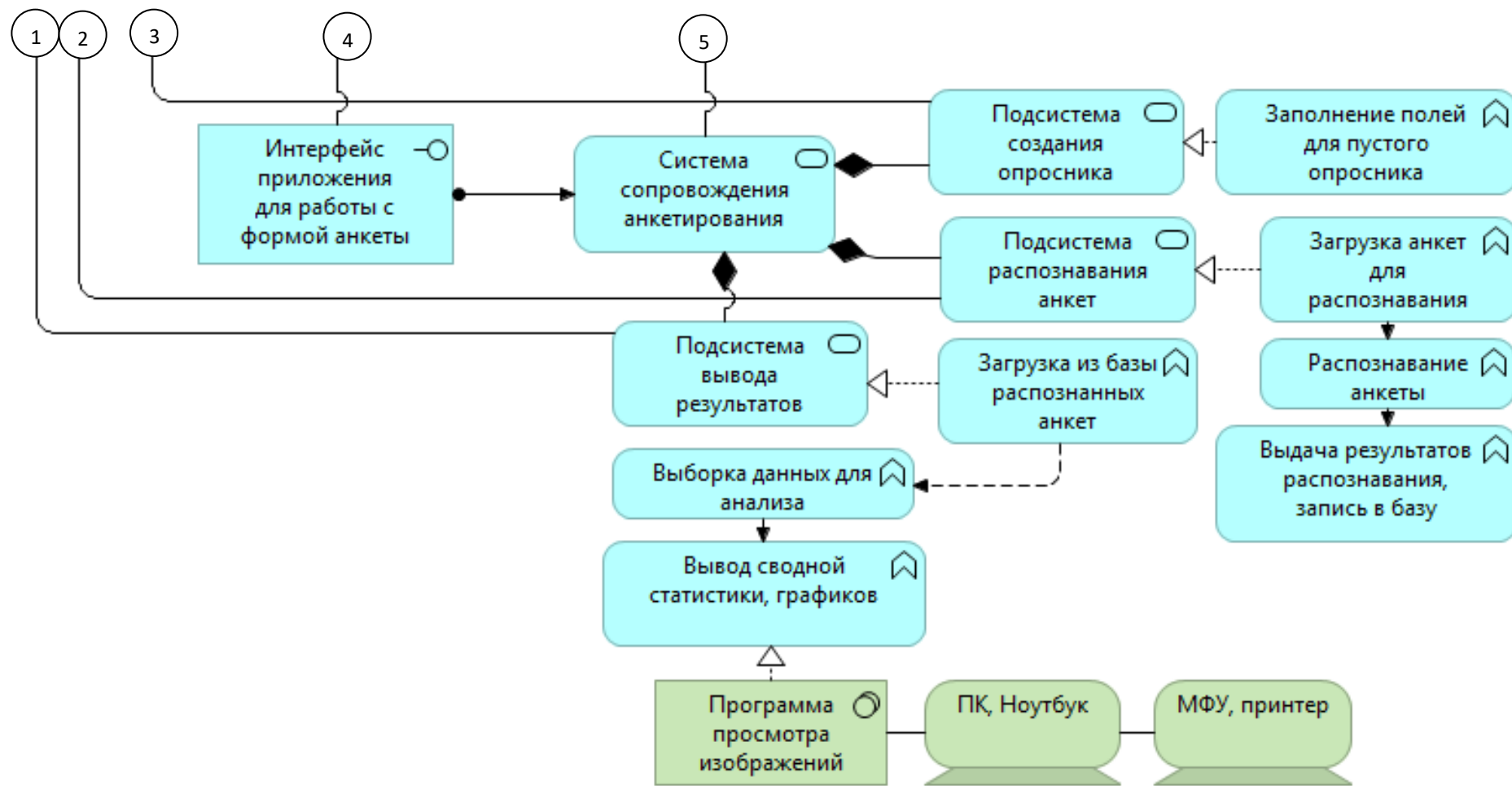


Рисунок 3.2. Бизнес слой системы

Система не автоматическая, а автоматизирована, т.е. требует вмешательства исследователя (человека). Роль человека в АС состоит в выборе методов и средств обработки данных, а также в принятии решений.

3.2. Программная реализация автоматизированной системы поддержки проведения и анализа результатов анкетирования.

Автоматизированная система поддержки проведения и анализа анкетных данных состоит из трех компонентов: приложение для создания шаблона опросника, консольное приложение для распознавания заполненных анкет, и консольное приложение для вывода статистических результатов собранных анкет.

3.2.1. Создание шаблона опросника

Приложение для создания шаблона опросника реализовано с помощью языка C# и интегрированной среды разработки программного обеспечения. C# – это объектно-ориентированный язык со строгой типизацией, позволяющий создавать различные безопасные и надежные приложения, работающие на платформе .NET Framework. Microsoft Visual Studio 2015 предоставляет собой полнофункциональную интегрированную среду разработки, удобные конструкторы пользовательского интерфейса, интегрированный отладчик и многие другие средства, которые упрощают разработку приложений на языке C# для платформы .NET Framework [27].

Основная цель написания программы для создания шаблона опросника, это возможность пользователю последовательно вводить вопросы и выбирать ответы к ним. Это позволяет сделать удобный интерфейс, написанный на Windows Forms. Выбор Windows Forms основывался на создании простого

приложения, не требовательного к ресурсам компьютера, с максимально понятным и функциональным интерфейсом.

Основная форма программы представлена на рисунке 3.3.

Рисунок 3.3. Основная форма программы.

Пользователь вводит название анкеты, и обращение к респонденту, если это необходимо. Для более удобной работы с анкетой создается отдельная директория, название которой совпадает с названием анкеты. Далее предусмотрено сохранение введенных данных в файл формата Excel. Для связи Microsoft Excel и среды Microsoft VisualStudio 2015 используется механизм COM (класс для работы с Excel). Кнопка далее рассчитана на работу как с новой анкетой, так и уже созданной.

На рисунке 3.4 показано сохранение файла с помощью открытия диалогового окна.

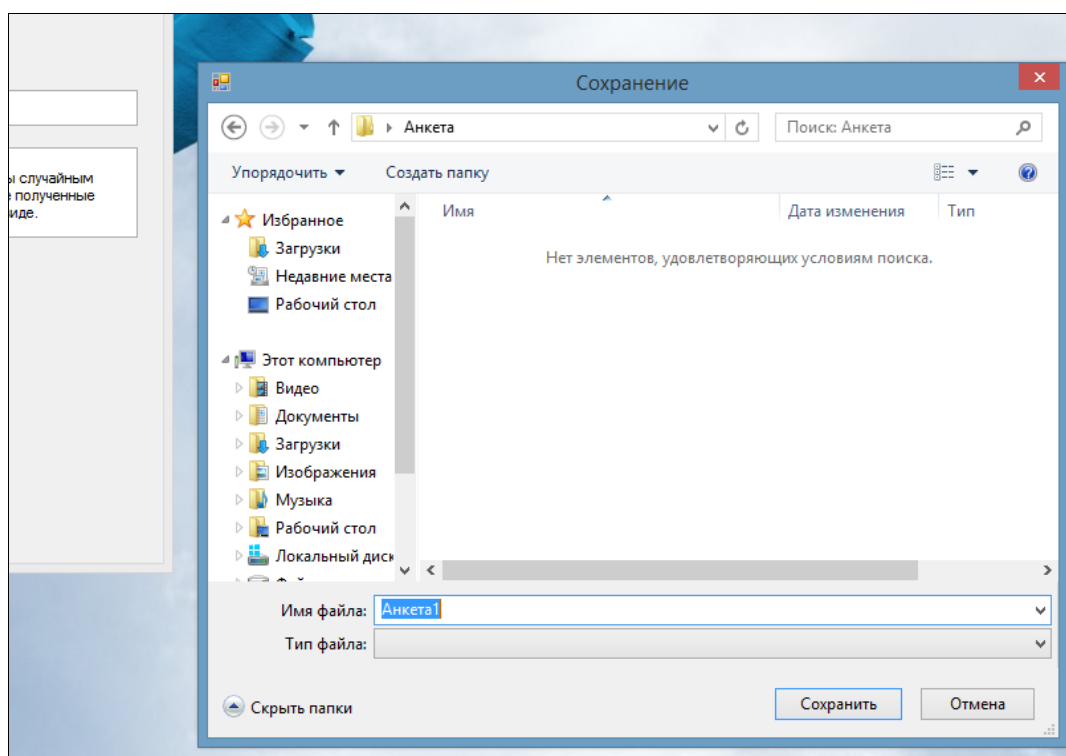


Рисунок 3.4. Диалоговое окно сохранения программы

После нажатия кнопки «далее», открывается вторая форма создания вопросов, состоящее из одного поля ввода вопроса, и открывающегося списка вариантов ответа. Предполагается, что вариант ответа должен быть один. Варианты ответа созданы стандартные, так и нестандартные (вводимые пользователем). Предусмотрено добавление вопроса как к новой анкете, так и к уже созданной.

На рис.3.5 представлена структура второй формы добавления вопроса и ответов к ним.

Создание анкеты. Шаг 2

Введите вопрос

Насколько часто вы пользуетесь образовательными ресурсами?

Варианты ответа

Стандартные варианты ответа

- да/нет
- да/нет/затрудняюсь ответить
- да/скорее да/скорее нет/нет
- правда/ложь
- редко/иногда/часто/очень часто
- никогда/очень редко/редко/иногда/очень часто/часто
- плохо/средне/хорошо/очень хорошо/отлично
- полностью согласен/полностью не согласен

записать в файл распечатать анкету

следующий вопрос предыдущий вопрос

выход

Рисунок 3.5. Создание вопроса к новой анкете.

Каждый созданный вопрос записывается в файл. Можно завершить создание анкеты или добавить еще один вопрос, используя навигацию.

На рис. 3.6 показана структура вопроса уже со своими вариантами ответа, где предусмотрено поле для ввода их количества. После введения в форму количества, создаются три динамических поля для ввода своих вариантов ответа.

The screenshot shows a window titled "Создание анкеты. Шаг 2" (Creating a survey. Step 2). The form contains the following elements:

- A text input field labeled "Введите вопрос" (Enter question) containing the text "1. Укажите Институт".
- A dropdown menu labeled "Варианты ответа" (Answer options) with the text "Свои варианты ответов(указать количе" (Your answer options (specify quantity)).
- A text input field labeled "Введите количество вариантов ответа" (Enter the number of answer options) containing the value "3".
- A "добавить" (add) button next to the quantity input.
- Three dynamically generated text input fields for answer options, containing the text "ИМиКН", "ФТИ", and "ИПиП".
- Five buttons at the bottom: "записать в файл" (save to file), "распечатать анкету" (print survey), "следующий вопрос" (next question), "предыдущий вопрос" (previous question), and "выход" (exit).

Рисунок 3.6. Вопрос со своими вариантами ответа.

Как только пользователь определится с количеством вопросов и ответов, созданную анкету можно перевести в машиночитаемый вид с помощью соответствующей кнопки. Далее формируется файл Word, реализуемый с помощью механизма COM (класс для работы с Word), с вводимыми данными, и таблицей с ячейками для распознавания маркеров.

Анкета для заполнения респондентом представлена на рис 3.7.

Анкета 1

Тюменский научно-образовательный центр Российской академии образования приглашает Вас принять участие в опросе, направленном на изучение способов взаимодействия студентов с образовательным контентом.

Анкетирование полностью анонимное, результаты будут обработаны с помощью методов математической статистики для выявления общего тренда.

1. Укажите Институт

| | |
|--------------------------|-------|
| <input type="checkbox"/> | ИМиКН |
| <input type="checkbox"/> | ФТИ |
| <input type="checkbox"/> | ИПИП |

2. Укажите ваш пол

| | |
|--------------------------|---|
| <input type="checkbox"/> | м |
| <input type="checkbox"/> | ж |

3. Укажите наиболее часто используемый образовательный контент

| | |
|--------------------------|---|
| <input type="checkbox"/> | Учебные материалы на бумажных носителях (Гуманитарные дисциплины) |
| <input type="checkbox"/> | Учебные материалы на бумажных носителях(ИТ-дисциплины) |
| <input type="checkbox"/> | Массовые открытые онлайн-курсы (Гуманитарные дисциплины) |
| <input type="checkbox"/> | Массовые открытые онлайн-курсы (ИТ-дисциплины) |

Рисунок 3.7. Анкета для заполнения респондентом.

Анкета создается в формате docs, где пользователь имеет возможность редактирования перед распечаткой.

На рис.3.8 вводимые данные показаны в формате Excel.

| id | text_q | n_answ | answ1 | answ2 | answ3 | answ4 |
|-----|--|--------|---|--|--|--|
| vr0 | 1. Укажите Институт | 3 | ИМиКН | ФТИ | ИПиП | |
| vr1 | 2. Укажите пол | 2 | м | ж | | |
| vr2 | 3. Укажите наиболее часто используемый образовательный контент | 4 | Учебные материалы на бумажных носителях (Гуманитарные дисциплины) | Учебные материалы на бумажных носителях(ИТ-дисциплины) | Массовые открытые онлайн-курсы (Гуманитарные дисциплины) | Массовые открытые онлайн-курсы (ИТ-дисциплины) |

Рисунок 3.8. Структура excel файла созданной анкеты

Программный код, соответствующий формам 1 (названия и обращения к респонденту) и 2 (создания вопроса и выбора вариантов ответа) представлены в Приложении 1 и Приложении 2.

3.2.2. Распознавание анкет

Для распознавания собранных анкет был выбран один из наиболее популярных и простых в реализации - метод распознавания образов на основе шаблонов [28]. Сама реализация представляет собой достаточно простые действия: сначала производится вычисление функции, показывающей степень совпадения шаблона с конечным изображением, затем поиск экстремума этой функции. В случае дипломной работы, одним из методов, требовались лишь знать центры закрасенных ячеек и попиксельное сравнение их. Если цвет пикселя центра ячейки шаблона не совпадает с цветом пикселя центра ячейки отмеченной анкеты, значит это и есть искомый ответ.

С практической точки зрения этот метод легко реализуемый, и многие коммерческие системы OCR (Optical Character Recognition) используют именно их. Но даже он имеет недостатки: даже небольшое темное пятнышко, попавшее на внешний контур символа, может существенно повлиять на

результат распознавания [29]. Поэтому для успешного распознавания анкет, респондентам настоятельно рекомендуется закрашивать полностью центр ячейки с выбранным вариантом ответа.

Алгоритм метода:

1. Предобработка метода (фильтрация, шумоподавление, бинаризация).
2. Определить вручную координаты центра шаблона и заполненной анкеты. Создать файл для записи массива распознанных ответов.
2. Распознавание (сопоставление центров ячеек с оригиналом).
3. Сравнение пикселей центров ячеек изображения с шаблоном:
 - 3.1. Если цвет пикселя совпадает, то соответствующий элемент массива принимает 0.
 - 3.2. Если цвет пикселя не совпадает, то соответствующий элемент массива принимает 1.
 - 3.3. После обнаружения несовпадающих ячеек цикл следует повторить.

Чаще всего именно предобработка данных является важным условием, повышающим эффективность распознавания искомых объектов.

Бинаризация является одним из самых распространенных методов предобработки изображения. Суть алгоритма заключается в сравнении яркости пикселя $V(x,y)$ с некоторым пороговым значением $V_t(x,y)$. Если яркость пикселя больше порогового значения, то цвет пикселя принимается равным 0, иначе 1.

Изображения, подлежащие распознаванию, могут быть зашумлены. Наиболее частая причина – пыль, попавшая в объектив или сканер. Поэтому необходима дополнительная предварительная фильтрация изображения.

На рисунке 3.9 показан пример шумоподавления изображений.

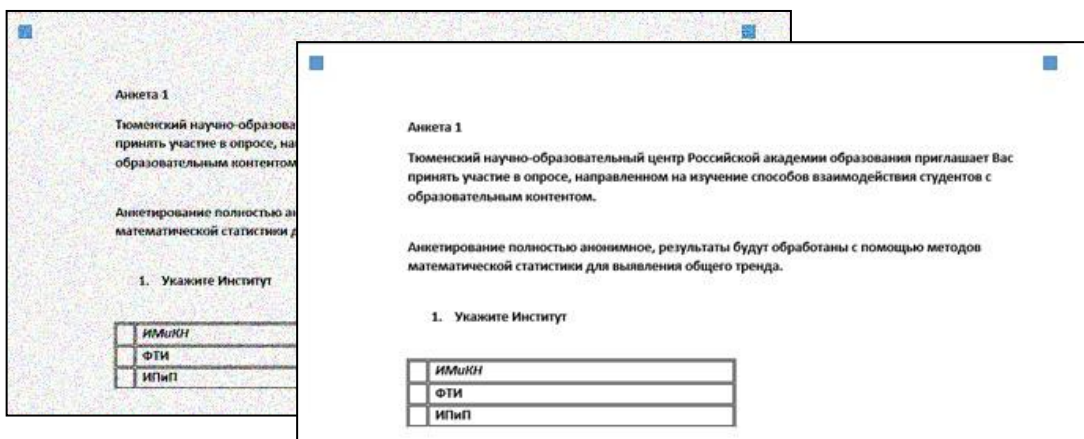


Рисунок 3.9. Пример шумоподавления изображения

Процесс сравнения отмеченной анкеты с шаблоном показан на рисунке 3.10.

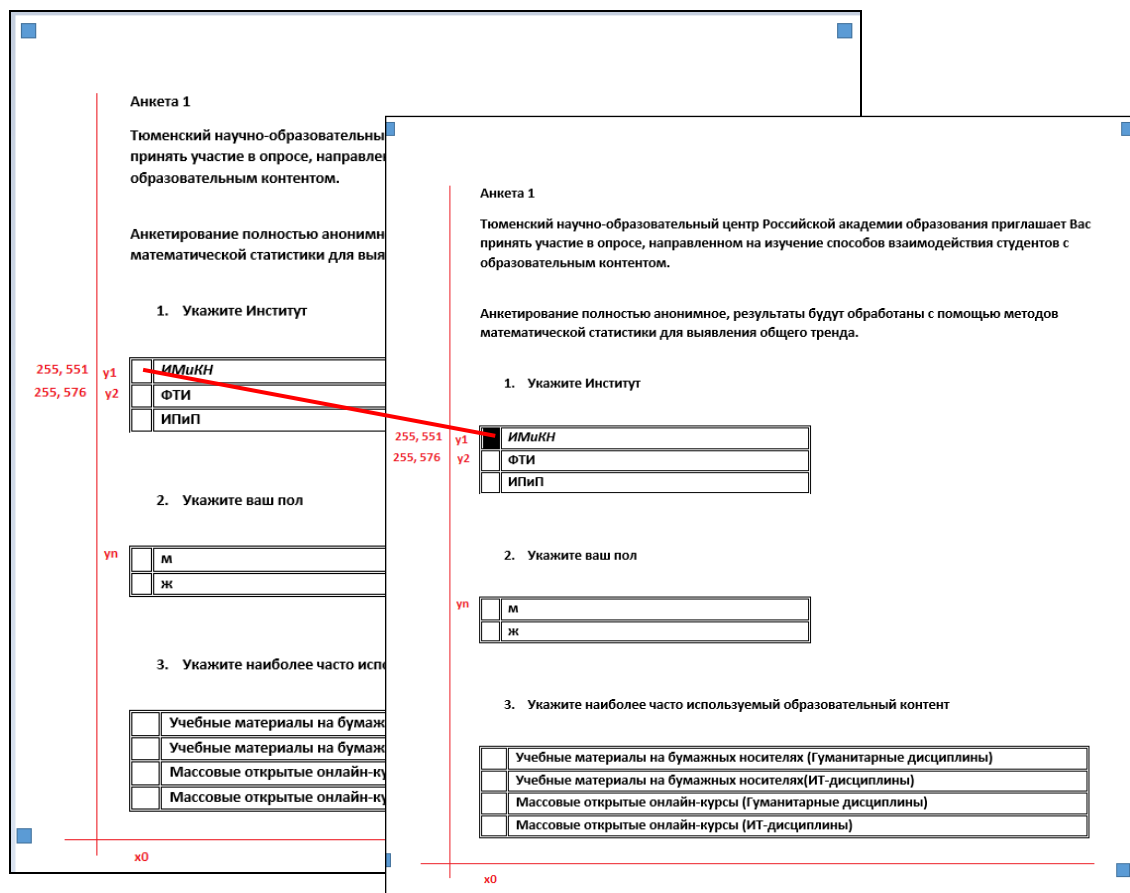


Рисунок 3.10 Процесс сравнения отмеченной анкеты с шаблоном

Другой вариант использования метода сравнения с шаблоном – сравнивать заполненную версию анкеты уже с заранее подготовленными отмеченными шаблонами. В этом случае, при совпадении цвета пикселей или признаков, в файл excel запишется 1, при несовпадении 0.

Процесс сравнения отмеченной анкеты с заранее подготовленными отмеченными шаблонами показан на рисунке 3.11.

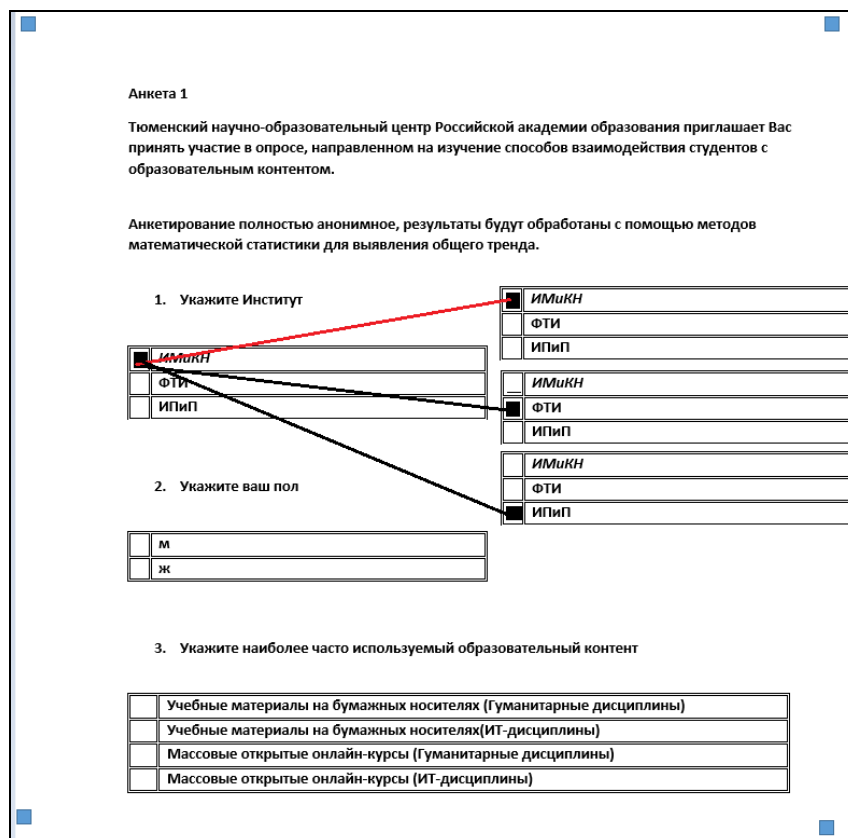


Рисунок 3.11 Процесс сравнения отмеченной анкеты с заранее подготовленными отмеченными шаблонами

Программная реализация осуществлена с помощью средств языка Python и дистрибутива Anaconda(Spyder).

На вход подается отсканированное изображение отмеченной анкеты(img1) и пустого шаблона(original):

```
img1 = cv2.imread("img.jpg")
img2 = cv2.imread("original.jpg")
```

Для работы с предобработкой используется библиотека `opencv`. (Open Source Computer Vision) – библиотека компьютерного зрения с открытым исходным кодом) – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом. Реализована на C/C++ [31].

Программный код предобработки изображения отсканированной анкеты:

```
ret,thresh1 = cv.threshold(img1,127,255,cv.THRESH_BINARY)#бинаризация  
blur = cv2.GaussianBlur(img1,(5,5),0) #удаление шумов с помощью гауссова размытия
```

Координаты центров ячеек отсканированной анкеты и шаблона:

```
f1 = img1[200, 577]  
f2 = img1[200, 621]  
f3 = img1[200, 653] #координаты центров ячеек отсканированной анкеты для первого  
вопроса(институты)  
  
f4 = img1[200, 885]  
f5 = img1[200,923] #координаты центров ячеек отсканированной анкеты для второго  
вопроса(пол)  
  
f6 = img1[200, 945]  
f7 = img1[200,964] #координаты центров ячеек отсканированной анкеты для третьего  
вопроса(образовательный контент)  
f8 = img1[200, 989]  
f9 = img1[200, 1007]  
  
f1_2 = img2[200, 577]  
f2_2 = img2[200, 621]  
f3_2 = img2[200, 653] #координаты центров ячеек шаблона для первого вопроса  
  
f4_2 = img2[200, 885]  
f5_2 = img2[200, 923] # координаты центров ячеек шаблона для второго вопроса  
  
f6_2 = img2[200,945]  
f7_2 = img2[200, 964]  
f8_2 = img2[200, 989] #координаты центров ячеек шаблона для первого вопроса  
f9_2 = img1[200, 1007]
```

Для, удобства созданы списки с центрами ячеек, для дальнейшего попарного сравнения шаблона с отсканированной анкетой. Сравнение происходит через генератор списков.

```
img_anketa = [f1,f2,f3,f4,f5,f6,f7,f8,f9]
img_original = [f1_2,f2_2,f3_2,f4_2,f5_2,f6_2,f7_2,f8_2,f9_2]
wb = xlwt.Workbook()
ws = wb.add_sheet('Answers')
[i for i, j in zip(img_anketa, img_original) ]
```

```
if i == j
    ws.write(0)
else:
    ws.write(1)
```

Результаты распознавания вопросов анкеты в формате Excel представлены на рисунке 3.12.

| | A | B | C | D |
|---|----|-----|------|------|
| 1 | id | ans | ans1 | ans2 |
| 2 | 0 | 0 | 1 | 0 |
| 3 | 1 | 0 | 1 | 0 |
| 4 | 2 | 0 | 0 | 1 |

Рисунок 3.12 Результаты распознавания вопросов анкеты

Для каждого вопроса анкеты создается отдельный файл Answers(n). На рисунке, представленном выше, представлен файл question0, где показан первый вопрос про институт и три варианта ответа, где два студента указали институт ФТИ, а третий ИПиП.

3.2.3. Анализ анкет

После этапа распознавания, у пользователя системы в директории (созданной еще на этапе создания опросника), есть в качестве входных данных

n-ое количество Excel анкет с ответами, пригодных для дальнейшего анализа результатов. Для дальнейшего анализа, все ответы на вопросы были объединены в один файл.

Обработка анкет осуществляется с помощью средств языка Python, библиотеки Pandas и дистрибутива Anaconda(Spyder).

Pandas это высокоуровневая Python библиотека для анализа данных. Чтобы эффективно работать с Pandas, необходимо освоить самые главные структуры данных библиотеки: DataFrame и Series. Без понимания, что они из себя представляют, невозможно в дальнейшем проводить качественный анализ. Структура/объект Series представляет из себя объект, похожий на одномерный массив, но отличительной его чертой является наличие ассоциированных меток, т.н. индексов, вдоль каждого элемента из списка. Такая особенность превращает его в ассоциативный массив или словарь в Python. DataFrame является табличной структурой данных. В любой таблице всегда присутствуют строки и столбцы. Столбцами в объекте DataFrame выступают объекты Series, строки которых являются их непосредственными элементами [URL: <https://pandas.pydata.org/>].

В Pandas предусмотрено слияние разных наборов данных. Сделать это можно с помощью функции merge(), (аналог join в SQL).

Визуализация ответов будет представлена на гистограммах, созданных с помощью библиотеки matplotlib.

Библиотека matplotlib - это библиотека двумерной графики для языка программирования python с помощью которой можно создавать высококачественные рисунки различных форматов. Matplotlib представляет собой модуль-пакет для python [URL: <https://matplotlib.org/>]

3.2.3.1. Анализ результатов

Пакет pandas позволяет собирать основные характеристики с помощью функции DataFrame.describe().

Без дополнительных параметров функция DataFrame.describe() предоставит следующую информацию для числовых типов данных, указанных в таблице 1:

Таблица 1. Функция DataFrame.describe()

| Вывод | Формулировка |
|-------|---|
| count | Подсчёт частоты того или иного события (сколько раз произошло событие). |
| mean | Среднее значение. |
| std | Стандартное отклонение (числовое значение, которое отображает изменение пределов данных). |
| min | Наименьшее число в наборе данных. |
| 25% | 25-й процентиль. |
| 50% | 50-й процентиль. |
| 75% | 75-й процентиль. |
| max | Максимальное число в наборе данных. |

Для дальнейшей работы с описательной статистикой были выполнены преобразования ответов. В случае с вопросом с институтом, ИМИКН имеет значение 0, ФТИ – 1, а ИПИП – 2.

Фрагмент кода преобразования значений столбцов

```
df = pd.read_excel(file_name)
df['ans2'] = df['ans2'].apply({1:2}.get)
```

Результат преобразования показан на рисунке 3.13.

| ans | ans1 | ans2 |
|-----|------|------|
| 0 | 1 | 0 |
| 0 | 1 | 0 |
| 0 | 0 | 2 |

Рисунок 3.13 Преобразование значений столбцов

Фрагмент кода статистического анализа для группы студентов:

```
qtext = ['Институт', 'Пол', 'Как часто']
# list of feature names
qkeys = ['q011', 'q012', 'q021']
# dict for processing
qdict = dict(zip(qkeys, qtext))
# students
df = pd.read_csv('DATA/students4nonzeros.csv')
writer_stud = pd.ExcelWriter('students_question3.xlsx')
df['department'] = df['department'].replace({0: 'ИМКН', 1: 'ФТИ', 2: 'ИПИП'})
df['pol'] = df['pol'].replace({0: 'm', 1: 'zh'})
# agregation for students
grouped = df.groupby(['pol', 'department'])

# processing
print ('-----STUDENTS-----')
for q in qkeys:
    print ('-----')
    print(qdict[q])
    descr = grouped[q].describe()
    x_mean = grouped[q].count()
    print(descr)
    # histograms
    x_mean.unstack().plot.bar(title = 'Сводная таблица', table = True)
    plt.xlabel("")
    plt.xticks([])
    #plt.show()
    # descr and hist to file
    name_out = 'out_stud'+q
```

```

descr.to_excel(writer_stud, sheet_name = q )
plt.savefig(name_out + '.png', format='png', dpi=100)
writer_stud.save()

```

Гистограмма с описательной статистикой представлена в соответствии с рисунком 3.13 и 3.14

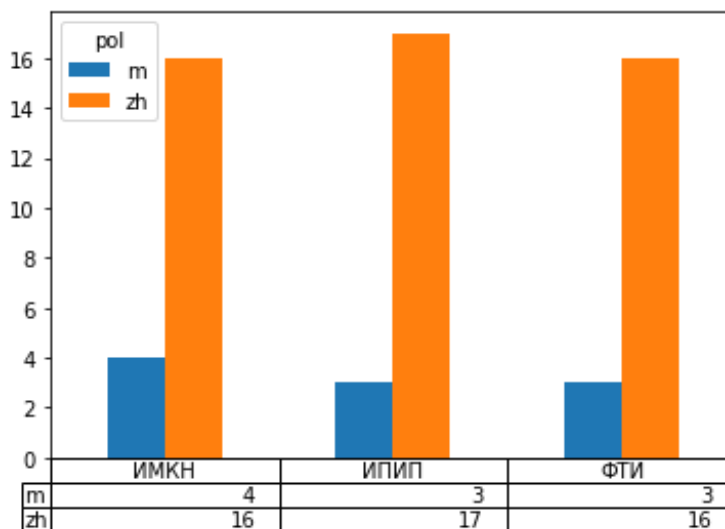


Рисунок 3.13 Сводные значения по институтам

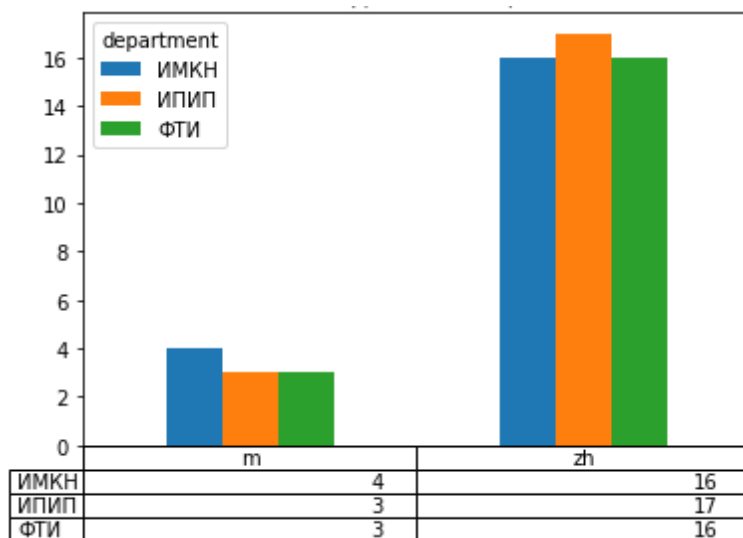


Рисунок 3.13 Сводные значения по полу

3.2.3.2. Кластерный анализ

Для работы с кластеризацией было выбрано исследование, проходившее в институтах ИМиКН и ИПиП. Исходный и преобразованный вид анкеты представлен в приложении 3. Для вариантов ответов заведен отдельный словарь, предоставленный в приложении 4.

Алгоритм реализован в библиотеке `scipy`, в модуле `cluster.vq` [URL: <http://scikit-learn.org/stable/modules/clustering.html>].

Пример кода кластеризации вариантов ответов для ИМиКН представлен ниже:

```
num_clusters = 4 #задается количество кластеров
km = KMeans(n_clusters=num_clusters, n_init=1, verbose=1, random_state=3)
#cluster for department=0
X0 = df1[df['department']==0]
clustered = km.fit(X0)# обучение
print("km.labels_.shape=%s" % km.labels_.shape)
# центры кластеров
#print("km.centers_=%s" % km.cluster_centers_)
df_centers = pd.DataFrame(km.cluster_centers_, index = ['cluster_' + str(i) for i in range(num_clusters)],
columns = qkeys)
df_centersT = df_centers.T
df_centersT['text'] = qtext
#print (df_centersT)
df_centersT.to_excel(writer_cluster, sheet_name = 'ИМКН - центры кластеров')
df_0 = pd.DataFrame(km.labels_, columns = ['cluster'])
df_0['year'] = df['year']
# agregation for students
grouped_0 = df_0.groupby(['cluster', 'year'])
name_out = 'out_stud3cluster_IMKN.csv'
gr_0 = grouped_0.size()
#gr_0.to_csv(name_out)
gr_0unstack = gr_0.unstack()
gr_0norm1 = gr_0unstack.div(gr_0unstack.sum(0).astype(float), axis = 1)
gr_0norm1T = gr_0norm1.T
print ('-----\n')
print ('Отдельная кластеризация (ИМКН): \n распределение курсов по кластерам абсолютное')
print (gr_0)
print (gr_0unstack)
```

```

print ('-----\n')
print ('Отдельная кластеризация (ИМКН): \n распределение курсов по кластерам относительное')
print (gr_0norm1T.aplymap(format))
print ('-----\n')
gr_0unstack.plot.bar(stacked=True, title = 'Отдельная кластеризация (ИМКН): \n распределение курсов
по кластерам абсолютное', table = True)
plt.xlabel("")
plt.xticks([])
gr_0norm1T.plot.bar(stacked=True, title = 'Отдельная кластеризация (ИМКН): \n распределение
курсов по кластерам относительное', table = True)
plt.xlabel("")
plt.xticks([])

```

В результате у нас появился список кластеров, каждый из которых содержит данные о входящих в него объектах, включая его имя и характеристики.

Значение центров кластеров по курсам представлен на рисунке 3.12.

| | cluster_0 | cluster_1 | cluster_2 | cluster_3 | text | | | | |
|------|-----------|-----------|-----------|-----------|--|--|--|--|--|
| q011 | 3,4375 | 2,674419 | 3,423077 | 1,689655 | 3.1.1. educational e-content: hum courses | | | | |
| q012 | 3,5625 | 3,604651 | 4,230769 | 4,206897 | 3.1.2. educational e-content: it courses | | | | |
| q021 | 3,0625 | 1,930233 | 3,576923 | 1,37931 | 3.2.1. educational p-content: hum courses | | | | |
| q022 | 2,854167 | 2,860465 | 3,038462 | 3,413793 | 3.2.2. educational p-content: it courses | | | | |
| q031 | 2,520833 | 1,860465 | 3,115385 | 1,517241 | 3.3.1. research e-content: hum courses | | | | |
| q032 | 3,145833 | 2,55814 | 3,769231 | 3,62069 | 3.3.2. research e-content: it courses | | | | |
| q041 | 2,104167 | 1,372093 | 2,730769 | 1,103448 | 3.4.1. research p-content: hum courses | | | | |
| q042 | 2,020833 | 1,697674 | 2,615385 | 2,655172 | 3.4.2. research p-content: it courses | | | | |
| q051 | 2,1875 | 1,55814 | 3,538462 | 1,241379 | 3.5.1. video-lectures: hum courses | | | | |
| q052 | 3,270833 | 2,860465 | 4,538462 | 3,931034 | 3.5.2. video-lectures: it courses | | | | |
| q061 | 1,645833 | 1,162791 | 3,115385 | 1,068966 | 3.6.1. audio-lectures: hum courses | | | | |
| q062 | 1,666667 | 1,325581 | 2,884615 | 2,172414 | 3.6.2. audio-lectures: it courses | | | | |
| q071 | 1,895833 | 1,372093 | 3,384615 | 1,206897 | 3.7.1. educational films: hum courses | | | | |
| q072 | 1,875 | 1,465116 | 3,307692 | 2,896552 | 3.7.2. educational films: it courses | | | | |
| q081 | 3,416667 | 1,604651 | 3,615385 | 1,103448 | 3.8.1. presentations: hum courses | | | | |
| q082 | 3,6875 | 2,488372 | 3,961538 | 3,517241 | 3.8.2. presentations: it courses | | | | |
| q091 | 3,625 | 1,906977 | 3,653846 | 1,275862 | 3.9.1. tests: hum courses | | | | |
| q092 | 3,791667 | 2,534884 | 4 | 3,241379 | 3.9.2. tests: it courses | | | | |
| q101 | 2,583333 | 1,488372 | 3,115385 | 1,206897 | 3.10.1. training or learning programm: hum courses | | | | |
| q102 | 3,208333 | 2,186047 | 4,153846 | 3,37931 | 3.10.2. training or learning programm: it courses | | | | |
| q111 | 2,75 | 1,27907 | 3,230769 | 1,137931 | 3.11.1. e-learning course: hum courses | | | | |
| q112 | 3,583333 | 2,44186 | 4,384615 | 4 | 3.11.2. e-learning course: it courses | | | | |
| q121 | 2,125 | 0,953488 | 3,615385 | 0,551724 | 3.12.1. MOOC: hum courses | | | | |
| q122 | 3 | 1,674419 | 4,615385 | 3,758621 | 3.12.2. MOOC: it courses | | | | |

Рисунок 3.12 Кластеризация по ИМиКН

Гистограмма распределения курсов по кластерам представлено в соответствии с рисунком 3.13.

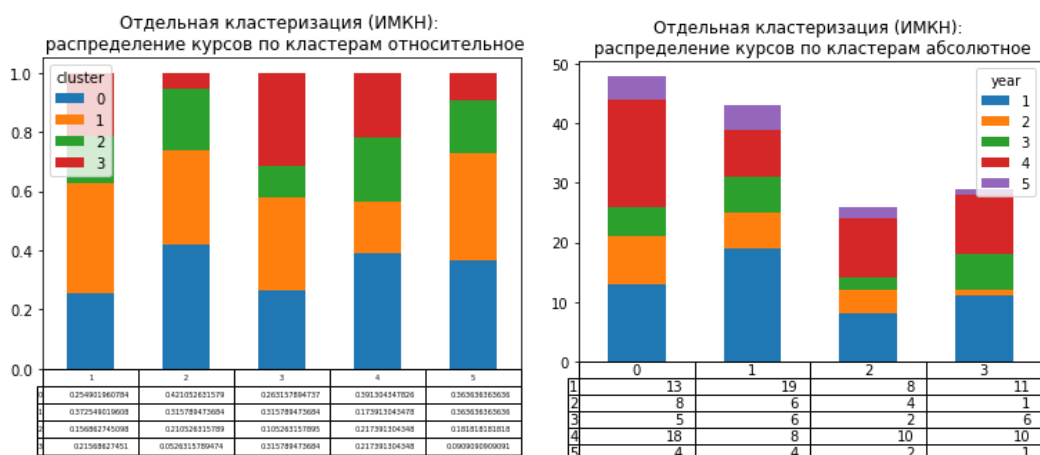


Рисунок 3.13. Распределение курсов по кластерам

Заключение

Все этапы проведения анкетирования, которые можно было автоматизировать, реализованы. С помощью подсистемы создания бумажного шаблона анкеты, пользователь сможет создать анкеты, пригодные для распознавания. Подсистема распознавания сможет уменьшить в разы время обработки отмеченных анкет, и сэкономит на человеческих ресурсах. Результаты проведения анкетирования можно увидеть наглядно, благодаря подсистеме анализа результатов.

Список литературы

1. Сравнительная оценка качества данных офлайн и онлайн-опросов [Электронный ресурс]. URL:<https://cyberleninka.ru/article/n/sravnitel'naya-otsenka-kachestva-dannyh-oflayni-onlayn-oprosov> (дата обращения: 22.04.2018 г.).
2. Социологические опросы в Интернете: возможности и ограничения [Электронный ресурс]. URL:<https://iq.hse.ru/more/sociology/sociologicheskie-oprosi-v-internete> (дата обращения: 22.04.2018)
3. Девятко И.Ф. Онлайн исследования и методология социальных наук: новые горизонты, новые (и не столь новые) трудности // Онлайн исследования в России 2.0: Сборник статей. М., 2010. С. 17–30.
4. Публикации РОС. [Электронный ресурс]. URL: http://www.ssa-rss.ru/index.php?page_id=40 (дата обращения: 03.10.2017 г.).
5. Российское общество социологов. [Электронный ресурс]. URL: <http://www.ssa-rss.ru/> (дата обращения: 03.10.2017 г.).
6. Шкурин Д.В. Технологии сбора и обработки данных по проекту «Российское студенчество о Великой Отечественной войне» // Война была позавчера... / Сб. статей Материалы мониторинга «Современное российское студенчество о Великой Отечественной войне». Екатеринбург, 2015. С. 84–89.
7. Маркович Д.Ж. Общая социология: Учебник. Изд.3-е перераб. и доп. /- М.: Гуманит. изд. центр ВЛАДОС, 2000.
8. Радугин А.А., Радугин К.А. Социология. М.: Центр, 2013. - 160с.
9. Лоу Д. Типы вопросов для опросников [Электронный ресурс]. URL: <http://books.irrp.org.ua/data-design/tipy-voprosov-dlya-oprosnikov/> (дата обращения: 23.04.2018 г.).
10. Вопросы для анкеты. Виды и примеры. [Электронный ресурс]. URL: <https://blog.anketolog.ru/2014/09/voprosy-dlja-ankety-vidy-i-primery/> (дата обращения: 11.05.2018 г.).
11. 5 сервисов для проведения онлайн опросов [Электронный ресурс] URL: <https://habr.com/post/297686/3/> (дата обращения: 26.04.2018 г.).

12. Какой сервис онлайн-опросов лучше? [Электронный ресурс]
URL:<https://blog.anketolog.ru/2016/08/servis-online-oprosov/><https://blog.anketolog.ru/2016/08/servis-online-oprosov/> (дата обращения: 15.05.2018 г.).
13. АБВУУ. Решения для интеллектуальной обработки информации. [Электронный ресурс] URL:<https://habrahabr.ru/company/abbyu/blog/312058/3/> (дата обращения: 26.02.2018 г.).
14. ECM-Journal. Основы электронного документооборота. [Электронный ресурс] URL:<https://ecm-journal.ru/post/Bumaga-vehlektronnom-dokumentoborote-ili-Obraznoe-myshlenie.aspx> (дата обращения: 21.02.2018 г.).
15. Создание машиночитаемых форм [Электронный ресурс] URL:
<http://docplayer.ru/257611-Sozdaniemashinochitaemyhform.html> (дата обращения: 23.03.2018 г.).
16. Термин «Машиночитаемость». [Электронный ресурс]. URL:
<https://ru.wikipedia.org/wiki/Машиночитаемость> (дата обращения: 22.02.2018 г.).
17. Работа с программой АБВУУ FormDesigner [Электронный ресурс]. URL: http://help.abbyu.cn/ruru/flexicapture/12/formdesigner/element_reference_mark (дата обращения: 15.05.2018 г.).
18. Ядов В.А. Социологическое исследование: методология программа методы/Учебники, монографии по социологии [Электронный ресурс]. URL: <http://socioline.ru/>
19. Мартышенко С.Н., Мартышенко Н.С., Кустов Д.А. «Совершенствование математического и программного обеспечения обработки первичных данных в экономических и социологических исследованиях», Известия Дальневосточного федерального университета. Экономика и управление, Вестник ТГЭУ №2, 2006 год.
20. Арлазаров В.В., Кривцов В.Е., Полевой Д.В., Слугин Д.Г., Янишевский И.М., «От графического образа к универсальному представлению формы документа»// Труды Института системного анализа Российской Академии наук. 2013г. Том 63.Выпуск 3. С.95-109.

21. Григорьев В.К., Грушин А.В., «Особенности распознавания информации об успеваемости и анкетировании студентов» «ВЕСТНИК МГТУ МИРЭА» № 4 2015 Том II.

22. Платонова И.Ю. Количественные и качественные методы в психолого-педагогическом исследовании. /Материал презентации [Электронный ресурс]. URL: <https://ppt-online.org/30248>

23. Бородкин Л.И. Математические модели в исторических исследованиях/ Портал ПостНаука <https://postnauka.ru/video/75717>

24. Кластерный анализ, Национальная библиотека им. Н. Э. Баумана Bauman National Library, [Электронный ресурс]. URL: <https://ru.bmstu.wiki/> (дата обращения: 02.06.2018 г.).

25. Systems Engineering Thinking Wiki [Электронный ресурс]. URL:<http://sewiki.ru/ArchiMate/> (дата обращения: 15.05.2018 г.).

26. Руководство по языку C#. [Электронный ресурс]. URL: <https://docs.microsoft.com/ru-ru/dotnet/csharp/> (дата обращения: 01.02.2018 г.).

27. Распознавание текста с использованием шаблонов <http://recog.ru/blog/ocr/9.html>

28. Предварительная обработка изображений, Национальная библиотека им. Н. Э. Баумана, Bauman National Library, [Электронный ресурс]. URL: <https://ru.bmstu.wiki/> (дата обращения: 02.06.2018 г.).

Приложение 1. Форма создания анкеты, директории работы с ней

```
namespace WindowsFormsApplication3
{
    public partial class Form2 : Form
    {
        Excel.Application oXL;
        Excel._Workbook oWB;
        Excel._Worksheet oSheet;
        object oEndOfDoc = "\\endofdoc";

        public Form2()
        {
            InitializeComponent();
        }
        SaveFileDialog dialog = new SaveFileDialog();
        private void button2_Click(object sender, EventArgs e)//Создание файла для работы с анкетой
        {
            Directory.CreateDirectory(textBox4.Text); )//Создание директории для работы с анкетами

            using (dialog)
            {
                if (dialog.ShowDialog() == System.Windows.Forms.DialogResult.OK)
                {
                    oXL = new Excel.Application();
                    oWB = (Excel._Workbook)(oXL.Workbooks.Add());
                    oSheet = (Excel._Worksheet)oWB.ActiveSheet; //Запись в созданный excel файл названия ячеек
                    oSheet.Cells[1, 1] = "id";
                    oSheet.Cells[1, 2] = "text_q";
                    oSheet.Cells[1, 3] = "answ1";
                    oSheet.Cells[1, 4] = "answ2";
                    oSheet.Cells[1, 5] = "answ3";
                    oSheet.Cells[1, 6] = "answ4";
                    oSheet.Cells[1, 7] = "answ5";
                    oSheet.Cells[1, 8] = "answ6";
                }
            }
            oXL.Visible = true;
            oXL.UserControl = true;
        }

        private void button1_Click(object sender, EventArgs e)//Переход на следующую форму
        {
            Form3 f3 = new Form3();
            f3.Show();
        }

        private void button6_Exit(object sender, EventArgs e)//Выход из приложения
        {
            Application.Exit();
        }
    }
}
```

Приложение 2. Форма записи вопроса и вариантов ответов

```
namespace WindowsFormsApplication3
{
    public partial class Form3 : Form
    {
        Word._Application oWord;
        Word._Document oDoc;
        object oEndOfDoc = "\\endofdoc";
        public Form3()
        {
            InitializeComponent();
        }

        private void button7_Exit(object sender, EventArgs e)//Выход из приложения
        {
            Application.Exit();
        }

        private void comboBox1_AnswerChanged(object sender, EventArgs e)//Изменения видимости
        comboBox1 в зависимости от выбора вариантов ответов
        {
            if (comboBox1.SelectedItem == "Стандартные варианты ответа")
            {
                radioButton1.Visible = true;
                radioButton2.Visible = true;
                radioButton3.Visible = true;
                radioButton4.Visible = true;
                radioButton5.Visible = true;
                radioButton6.Visible = true;
                radioButton7.Visible = true;
                radioButton8.Visible = true;
                textBox2.Visible = false;
                label4.Visible = false;
                button1.Visible = false;
            }

            if (comboBox1.SelectedItem == "Свои варианты ответов(указать количество)")
            {
                textBox2.Visible = true;
                label4.Visible = true;
                radioButton1.Visible = false;
                radioButton2.Visible = false;
                radioButton3.Visible = false;
                radioButton4.Visible = false;
                radioButton5.Visible = false;
                radioButton6.Visible = false;
                radioButton7.Visible = false;
                radioButton8.Visible = false;
                button1.Visible = true;
            }
        }

        List<TextBox> MyVariantsTextbox = new List<TextBox>();

        private void button1_ClickToNumberVariants(object sender, EventArgs e) //создание
        динамического текстового для своих вариантов ответов
        {
            for (int i = 0, x = 1; i < int.Parse(textBox2.Text); i++, x++)
            {
                MyVariantsTextbox.Add(new TextBox());
                MyVariantsTextbox[i].Location = new System.Drawing.Point(240, (190 + i * 30));
                this.Controls.Add(MyVariantsTextbox[i]);
            }
        }
    }
}
```

```

private void button2_WriteToFileClick(object sender, EventArgs e)//запись анкеты в файл excel
{
    OpenFileDialog dialog = new OpenFileDialog();
    using (dialog)
    {
        if (dialog.ShowDialog() == System.Windows.Forms.DialogResult.OK)
        {
            Excel.Application Excel_Book = new Excel.Application();
            Excel.Worksheet oSheet;
            Excel_Book.Workbooks.Open(dialog.FileName);
            Excel_Book.Cells[2, 2] = textBox1.Text;

            if (comboBox1.SelectedItem == "Стандартные варианты ответа")
            {
                if (radioButton1.Checked)
                {
                    Excel_Book.Cells[1, 3] = "n_answ";
                    Excel_Book.Cells[2, 3] = "2";
                    Excel_Book.Cells[2, 4] = "да";
                    Excel_Book.Cells[2, 5] = "нет";
                }

                if (radioButton6.Checked)
                {
                    Excel_Book.Cells[1, 3] = "n_answ";
                    Excel_Book.Cells[2, 3] = "3";
                    Excel_Book.Cells[2, 4] = "да";
                    Excel_Book.Cells[2, 5] = "нет";
                    Excel_Book.Cells[2, 6] = "затрудняюсь ответить";
                }

                if (radioButton2.Checked)
                {
                    Excel_Book.Cells[1, 3] = "n_answ";
                    Excel_Book.Cells[2, 3] = "4";
                    Excel_Book.Cells[2, 4] = "да";
                    Excel_Book.Cells[2, 5] = "скорее да";
                    Excel_Book.Cells[2, 6] = "скорее нет";
                    Excel_Book.Cells[2, 7] = "нет";
                }

                if (radioButton3.Checked)
                {
                    Excel_Book.Cells[1, 3] = "n_answ";
                    Excel_Book.Cells[2, 3] = "2";
                    Excel_Book.Cells[2, 4] = "правда";
                    Excel_Book.Cells[2, 5] = "ложь";
                }

                if (radioButton5.Checked)
                {
                    Excel_Book.Cells[1, 3] = "n_answ";
                    Excel_Book.Cells[2, 3] = "3";
                    Excel_Book.Cells[2, 4] = "никогда";
                    Excel_Book.Cells[2, 5] = "иногда";
                    Excel_Book.Cells[2, 6] = "часто";
                }

                if (radioButton4.Checked)
                {
                    Excel_Book.Cells[1, 3] = "n_answ";
                    Excel_Book.Cells[2, 3] = "5";
                    Excel_Book.Cells[2, 4] = "никогда";
                    Excel_Book.Cells[2, 5] = "очень редко";
                    Excel_Book.Cells[2, 6] = "иногда";
                }
            }
        }
    }
}

```

```

Excel_Book.Cells[2, 7] = "достаточно часто";
Excel_Book.Cells[2, 8] = "регулярно";
}

if (radioButton7.Checked)
{
Excel_Book.Cells[1, 3] = "n_answ";
Excel_Book.Cells[2, 3] = "5";
Excel_Book.Cells[2, 4] = "плохо";
Excel_Book.Cells[2, 5] = "средне";
Excel_Book.Cells[2, 6] = "хорошо";
Excel_Book.Cells[2, 7] = "очень хорошо";
Excel_Book.Cells[2, 8] = "отлично";
}

if (radioButton8.Checked)
{
Excel_Book.Cells[1, 3] = "n_answ";
Excel_Book.Cells[2, 3] = "2";
Excel_Book.Cells[2, 4] = "полностью согласен";
Excel_Book.Cells[2, 5] = "полностью не согласен";
}
Excel_Book.Save(dialog.FileName);
}

if (comboBox1.SelectedItem == "Свои варианты ответов(указать количество)")
{
for (int i = 0; i < MyVariantsTextbox.Count; i++)
{
Excel_Book.Cells[i, i] = MyVariantsTextbox[i].Text;
}
}
Excel_Book.Save(dialog.FileName);
}
}
}

private void button5_AnketForm(object sender, EventArgs e)//вывод анкеты из файла excel в
файл word
{
OpenFileDialog dialog = new OpenFileDialog();
using (dialog)
{
if (dialog.ShowDialog() == DialogResult.OK)
{
Excel.Application Excel_Book = new Excel.Application();
Excel_Book.Workbooks.Open(dialog.FileName);
oWord = new Word.Application();
oDoc = oWord.Documents.Add();
Word.Paragraph oPara1;
oPara1 = oDoc.Content.Paragraphs.Add();
oPara1.Range.Font.Bold = 1;
oPara1.Format.SpaceAfter = 24;
oPara1.Range.InsertParagraphAfter();
Excel_Book.Range["B2", "B6"].Copy();
oDoc.Range().PasteSpecial();

}

if (radioButton1.Checked)//если да/нет
{

Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 2, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 1;
oTable.Cell(1, 2).Range.Text = "да";
oTable.Cell(1, 1).Height = 15;
}
}
}
}

```

```

oTable.Cell(1, 1).Width= 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;
oTable.Cell(2, 2).Range.Text = "нет";

oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.InsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}

if (radioButton6.Checked)//если никогда/иногда/часто
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 3, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 3;

oTable.Cell(1, 2).Range.Text = "да";
oTable.Cell(2, 2).Range.Text = "нет";
oTable.Cell(3, 2).Range.Text = "затрудняюсь ответить";
oTable.Cell(1, 1).Height = 15;
oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;
oTable.Cell(3, 1).Height = 15;
oTable.Cell(3, 1).Width = 15;
oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.InsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}

if (radioButton2.Checked)
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 4, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 3;
oTable.Cell(1, 2).Range.Text = "да";
oTable.Cell(2, 2).Range.Text = "скорее да";
oTable.Cell(3, 2).Range.Text = "скорее нет";
oTable.Cell(4, 2).Range.Text = "нет";
oTable.Cell(1, 1).Height = 15;
oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;
oTable.Cell(3, 1).Height = 15;
oTable.Cell(3, 1).Width = 15;
oTable.Cell(4, 1).Height = 15;
oTable.Cell(4, 1).Width = 15;
oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.InsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}
if (radioButton3.Checked)
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 2, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 3;
oTable.Cell(1, 2).Range.Text = "правда";
oTable.Cell(2, 2).Range.Text = "ложь";
oTable.Cell(1, 1).Height = 15;

```

```

oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;

oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}
if (radioButton5.Checked)
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 3, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 3;
oTable.Cell(1, 2).Range.Text = "никогда";
oTable.Cell(2, 2).Range.Text = "иногда";
oTable.Cell(3, 2).Range.Text = "часто";
oTable.Cell(1, 1).Height = 15;
oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;
oTable.Cell(3, 1).Height = 15;
oTable.Cell(3, 1).Width = 15;
oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}
if (radioButton4.Checked)
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 6, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 3;
oTable.Cell(1, 2).Range.Text = "никогда";
oTable.Cell(2, 2).Range.Text = "очень редко";
oTable.Cell(3, 2).Range.Text = "редко";
oTable.Cell(4, 2).Range.Text = "иногда";
oTable.Cell(5, 2).Range.Text = "очень часто";
oTable.Cell(6, 2).Range.Text = "часто";

oTable.Cell(1, 1).Height = 15;
oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;
oTable.Cell(3, 1).Height = 15;
oTable.Cell(3, 1).Width = 15;
oTable.Cell(4, 1).Height = 15;
oTable.Cell(4, 1).Width = 15;
oTable.Cell(5, 1).Height = 15;
oTable.Cell(5, 1).Width = 15;
oTable.Cell(6, 1).Height = 15;
oTable.Cell(6, 1).Width = 15;

oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}
if (radioButton7.Checked)//если никогда/иногда/часто
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 5, 2);

```



```

oTable.Range.ParagraphFormat.SpaceAfter = 3;

oTable.Cell(1, 2).Range.Text = "плохо";
oTable.Cell(2, 2).Range.Text = "средне";
oTable.Cell(3, 2).Range.Text = "хорошо";
oTable.Cell(4, 2).Range.Text = "очень хорошо";
oTable.Cell(5, 2).Range.Text = "отлично";

oTable.Cell(1, 1).Height = 15;
oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;
oTable.Cell(3, 1).Height = 15;
oTable.Cell(3, 1).Width = 15;
oTable.Cell(4, 1).Height = 15;
oTable.Cell(4, 1).Width = 15;
oTable.Cell(5, 1).Height = 15;
oTable.Cell(5, 1).Width = 15;

oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}
if (radioButton8.Checked)//если никогда/иногда/часто
{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;
oTable = oDoc.Tables.Add(wrdRng, 2, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 3;

oTable.Cell(1, 2).Range.Text = "полностью согласен";
oTable.Cell(2, 2).Range.Text = "полностью не согласен";
oTable.Cell(1, 1).Height = 15;
oTable.Cell(1, 1).Width = 15;
oTable.Cell(2, 1).Height = 15;
oTable.Cell(2, 1).Width = 15;

oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oWord.Visible = true;
}

else//если свои варианты ответа

{
Word.Table oTable;
Word.Range wrdRng = oDoc.Bookmarks.get_Item(ref oEndOfDoc).Range;

oTable = oDoc.Tables.Add(wrdRng, 6, 2);
oTable.Range.ParagraphFormat.SpaceAfter = 1;

for (int i = 0; i < MyVariantsTextbox.Count; i++)
oTable.Cell(i, i).Range.Text = MyVariantsTextbox[i].Text;
oTable.Rows[1].Range.Font.Bold = 1;
oTable.Rows[1].Range.Font.Italic = 1;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
oTable.Borders.OutsideLineStyle = Word.WdLineStyle.wdLineStyleDouble;
}

oDoc.Shapes.AddShape((int)Microsoft.Office.Core.MsoAutoShapeType.msoShapeRectangle, 10, 10,
10, 10); //реперные точки
oDoc.Shapes.AddShape((int)Microsoft.Office.Core.MsoAutoShapeType.msoShapeRectangle, 570, 10,
10, 10);

```

```

    oDoc.Shapes.AddShape((int)Microsoft.Office.Core.MsoAutoShapeType.msoShapeRectangle, 10, 800,
10, 10);
    oDoc.Shapes.AddShape((int)Microsoft.Office.Core.MsoAutoShapeType.msoShapeRectangle, 570, 800,
10, 10);
    }
    oWord.Visible = true;
    }

    private void button6_Back(object sender, EventArgs e)
    {
        Form2 f2 = new Form2();
        f2.Show();
    }

    private void button3_NextQuestions(object sender, EventArgs e)
    {
        textBox1.Clear();
    }

    int index = 0;
    Excel.Worksheet oSheet;
    private void button4_PreviousQuestions(object sender, EventArgs e)
    {
        index -= 1;
        textBox1.Text = oSheet.Cells[index];
    }
}

```

Приложение 3. Исходная анкета для кластеризации

Тюменский научно-образовательный центр Российской академии образования приглашает Вас принять участие в опросе, направленном на изучение способов взаимодействия студентов с образовательным контентом.

Анкетирование полностью анонимное, результаты будут обработаны с помощью методов математической статистики для выявления общего тренда.

1. Укажите свой Институт: 2. Вы обучаетесь:

| | |
|-------|--|
| ИМиКН | |
| ИПиП | |

| | |
|------------------|--|
| Бак/спец. 1 курс | |
| Бак/спец. 2 курс | |
| Бак/спец. 3 курс | |
| Бак/спец. 4 курс | |
| Магистратура | |

3. Как часто, по Вашему мнению, студенты используют разные виды образовательного контента:

| ОБРАЗОВАТЕЛЬНЫЙ КОНТЕНТ | | | никогда | очень редко | иногда | достаточно часто | регулярно |
|-------------------------|--|-------------------------|---------|-------------|--------|------------------|-----------|
| 1.1 | Электронные учебные материалы: учебники, учебные пособия, учебно-методические пособия, практикумы, самоучители, конспекты лекций | Гуманитарные дисциплины | | | | | |
| 1.2 | | ИТ - дисциплины | | | | | |
| 2.1 | Учебные материалы на бумажных носителях: учебники, учебные пособия, учебно-методические пособия, практикумы, самоучители, конспекты лекций | Гуманитарные дисциплины | | | | | |
| 2.2 | | ИТ - дисциплины | | | | | |
| 3.1 | Электронные научные и/или технические материалы: монографии, статьи, отчеты, руководства | Гуманитарные дисциплины | | | | | |
| 3.2 | | ИТ - дисциплины | | | | | |
| 4.1 | Научные и/или технические материалы на бумажных носителях: монографии, статьи, отчеты, руководства | Гуманитарные дисциплины | | | | | |
| 4.2 | | ИТ - дисциплины | | | | | |
| 5.1 | Видеолекции | Гуманитарные дисциплины | | | | | |
| 5.2 | | ИТ - дисциплины | | | | | |
| 6.1 | Аудиолекции | Гуманитарные дисциплины | | | | | |
| 6.2 | | ИТ - дисциплины | | | | | |
| 7.1 | Учебные фильмы | Гуманитарные дисциплины | | | | | |
| 7.2 | | ИТ - дисциплины | | | | | |
| 8.1 | Презентации | Гуманитарные дисциплины | | | | | |
| 8.2 | | ИТ - дисциплины | | | | | |
| 9.1 | Тестовые задания | Гуманитарные дисциплины | | | | | |
| 9.2 | | ИТ - дисциплины | | | | | |
| 10.1 | Электронные тренажеры (компьютерные обучающие программы) | Гуманитарные дисциплины | | | | | |
| 10.2 | | ИТ - дисциплины | | | | | |
| 11.1 | Электронные учебные курсы | Гуманитарные дисциплины | | | | | |
| 11.2 | | ИТ - дисциплины | | | | | |
| 12.1 | Массовые открытые онлайн-курсы (включают видеолекции, тестирование и др.) | Гуманитарные дисциплины | | | | | |
| 12.2 | | ИТ - дисциплины | | | | | |

Фрагмент анкеты, приведенный под автоматизированную систему:

Анкета 1

Тюменский научно-образовательный центр Российской академии образования приглашает Вас принять участие в опросе, направленном на изучение способов взаимодействия студентов с образовательным контентом.

Анкетирование полностью анонимное, результаты будут обработаны с помощью методов математической статистики для выявления общего тренда.

1. Укажите Институт

| | |
|--------------------------|-------|
| <input type="checkbox"/> | ИМиКН |
| <input type="checkbox"/> | ИПиП |

2. Курс обучения

| | |
|--------------------------|-----------------|
| <input type="checkbox"/> | Бак/спец. 1курс |
| <input type="checkbox"/> | Бак/спец. 2курс |
| <input type="checkbox"/> | Бак/спец. 3курс |
| <input type="checkbox"/> | Бак/спец. 4курс |
| <input type="checkbox"/> | Магистратура |

3. Как часто вы используете учебные материалы на бумажных носителях для гуманитарных дисциплин?

| | |
|--------------------------|------------------|
| <input type="checkbox"/> | никогда |
| <input type="checkbox"/> | очень редко |
| <input type="checkbox"/> | иногда |
| <input type="checkbox"/> | достаточно часто |
| <input type="checkbox"/> | регулярно |

4. Как часто вы используете учебные материалы на бумажных носителях для IT дисциплин?

| | |
|--------------------------|------------------|
| <input type="checkbox"/> | никогда |
| <input type="checkbox"/> | очень редко |
| <input type="checkbox"/> | иногда |
| <input type="checkbox"/> | достаточно часто |
| <input type="checkbox"/> | регулярно |

Приложение 4. Словарь вариантов ответов для кластерного анализа

q011: 3.1.1. educational e-content: hum courses
q012: 3.1.2. educational e-content: it courses
q021: 3.2.1. educational p-content: hum courses
q022: 3.2.2. educational p-content: it courses
q031: 3.3.1. research e-content: hum courses
q032: 3.3.2. research e-content: it courses
q041: 3.4.1. research p-content: hum courses
q042: 3.4.2. research p-content: it courses
q051: 3.5.1. video-lectures: hum courses
q052: 3.5.2. video-lectures: it courses
q061: 3.6.1. audio-lectures: hum courses
q062: 3.6.2. audio-lectures: it courses
q071: 3.7.1. educational films: hum courses
q072: 3.7.2. educational films: it courses
q081: 3.8.1. presentations: hum courses
q082: 3.8.2. presentations: it courses
q091: 3.9.1. tests: hum courses
q092: 3.9.2. tests: it courses
q101: 3.10.1. training or learning programmes: hum courses
q102: 3.10.2. training or learning programmes: it courses
q111: 3.11.1. e-learning course: hum courses
q112: 3.11.2. e-learning course: it courses
q121: 3.12.1. MOOC: hum courses
q122: 3.12.2. MOOC: it courses