

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт кибернетики
Направление подготовки Прикладная математика и информатика
Кафедра прикладной математики

БАКАЛАВРСКАЯ РАБОТА

**Разработка алгоритмического и программного обеспечения для сбора, обработки и
анализа социологических данных**

УДК _____

Студент

8621	Ковригина Т.С.			
Группа	ФИО		Подпись	Дата

Руководитель

Профессор	Берестнева О.Г.	Д.Т.Н.		
Должность	ФИО	Ученая степень, звание	Подпись	Дата

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Ассистент	Николаенко Валентин Сергеевич			
Должность	ФИО	Ученая степень, звание	Подпись	Дата

По разделу «Социальная ответственность»

Доцент	Антоневич Ольга Алексеевна	К.Т.Н		
Должность	ФИО	Ученая степень, звание	Подпись	Дата

ДОПУСТИТЬ К ЗАЩИТЕ:

Зав. кафедрой	Гергет Ольга Михайловна	К.Т.Н.		
Должность	ФИО	Ученая степень, звание	Подпись	Дата

Томск – 2016 г.

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт кибернетики
Направление подготовки Прикладная математика и информатика
Кафедра прикладной математики

УТВЕРЖДАЮ:
Зав. кафедрой

(Подпись)

(Дата)

(Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

бакалаврской работы

Студенту:

Группа	ФИО
8Б21	Ковригиной Таисии Сергеевны

Тема работы:

Разработка алгоритмического и программного обеспечения для сбора, обработки и анализа социологических данных

Утверждена приказом директора Института
кибернетики (дата, номер)

Срок сдачи студентом выполненной работы: (дата)

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Документация проекта РГНФ №15-03-00366; Типовой инструментарий; Данные социологических исследований по Томской области в 2015 году;
---------------------------------	--

Перечень подлежащих исследованию, проектированию и разработке вопросов		Введение 1. Постановка задачи 2. Структура разрабатываемого сайта 3. Выбор и характеристика методов анализа экспериментальных данных 4. Программная реализация	
Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>		Слайды презентации;	
Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>			
Раздел		Консультант	
1. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение		Николаенко Валентин Сергеевич, ассистент	
2. Социальная ответственность		Антоневич Ольга Алексеевна, к.т.н., доцент	
3. Английский язык			
Названия разделов, которые должны быть написаны на русском и иностранном языках:			

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата

профессор	Берестнева Ольга Григорьевна	д.т.н.		
-----------	---------------------------------	--------	--	--

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8Б21	Ковригина Таисия Сергеевна		

РЕФЕРАТ

Выпускная квалификационная работа бакалавра выполнена на 64 листах машинописного текста, 2 рисунка, 14 таблиц, 20 источников, 2 приложения.

Ключевые слова: сайт, алгоритм, кластерный анализ, база данных, data mining.

Работа посвящена разработке сайта, предназначенного для проведения социальных и психологических исследований.

Пояснительная записка состоит из четырёх разделов.

В первом разделе представлена постановка задачи.

Во втором разделе приведены требования к разрабатываемому сайту и представлены основные модели поведения пользователей с различными правами доступа и очерчены границы их возможностей.

Третий раздел посвящён рассмотрению методов обработки получаемых данных.

Четвёртый раздел отводится для оценки и выбора программных средств, позволяющих реализовать заявленный функционал сайта.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	7
ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ, СОКРАЩЕНИЯ, НОРМАТИВНЫЕ ССЫЛКИ.....	9
1 ОБЗОР ЛИТЕРАТУРЫ.....	11
2 ПОСТАНОВКА ЗАДАЧИ.....	12
2.1 Характеристика экспериментальных данных	12
3 СТРУКТУРА САЙТА.....	14
3.1 Технические требования к подсистеме.....	14
3.2 Модели поведения пользователей	15
3.2.1 Сценарии поведения администратора	15
3.2.3 Сценарии поведения пользователя «Гость»	18
4 МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ.....	20
4.1 Выбор методов обработки данных	21
4.2 Методы Data Mining.....	23
4.2.1 Методы кластерного анализа.....	26
5 ПРОГРАММНЫЕ СРЕДСТВА.....	32
5.1 Оценка программных средств Data Mining	32
6 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ	34
7 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	43
ЗАКЛЮЧЕНИЕ	57
СПИСОК ПУБЛИКАЦИЙ	58
СПИСОК ИСТОЧНИКОВ	59
ПРИЛОЖЕНИЯ	62
Приложение А.1	62
Приложение А.2	63

ВВЕДЕНИЕ

Важную роль в обеспечении благополучия человека и общества играют психологические исследования и службы, в том числе, проводимые на базе высших учебных заведений. Одним из средств проведения тестирований являются сетевые ресурсы, как узко специализированные, так и многофункциональные. Так или иначе, помимо сбора информации, данные ресурсы должны предоставлять возможность накопления данных об участниках, результатах и процессе прохождения тестов, а так же обработки имеющейся информации.

Портал позволит оперативно вносить корректировки в тестирующие приложения, извлекать из базы данных информацию о конкретном тестирующемся, но, к сожалению, пока не предоставляет простого в обращении набора инструментов для обработки больших объёмов накопленных результатов. Необходимо создать портал, дающий аналитикам и психологам возможность работы с информацией без использования навыков программирования (которые психолог или математик иметь не обязан) и, помимо этого, расширяющий возможности обратной связи и взаимодействия психолога собственно с коллективом программистов, разработавших и сопровождающих данный ресурс.

Целью работы является разработка многофункционального сайта, для организации психологического on-line тестирования для различных групп пользователей, в том числе абитуриентов ТПУ.

Для достижения поставленной цели необходимо решить ряд задач:

- определить желаемую функциональность портала и границы полномочий пользователей с различными правами доступа;
- изучить методы многомерного статистического анализа;
- выбрать среду для реализации;
- разработать и реализовать портал;

- провести тестирование портала и оценить возможности обработки имеющихся данных;
- на основе результатов тестирования сформулировать выводы об эффективности реализации портала, дальнейших возможностях её развития и методах устранения возможных проблем, выявленных в ходе выполнения проекта.

ОПРЕДЕЛЕНИЯ, СОКРАЩЕНИЯ, ОБОЗНАЧЕНИЯ, НОРМАТИВНЫЕ ССЫЛКИ

В данной работе применены следующие термины и аббревиатуры с соответствующими определениями:

data mining – специальное название, которое используется для обозначения совокупности методов обнаружения в данных ранее неизвестных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности;

HTTP-сервер - сервер, принимающий HTTP-запросы от клиентов, обычно веб-браузеров, и выдающий им HTTP-ответы, как правило, вместе с HTML-страницей, изображением, файлом, медиа-поток или другими данными;

Кластер - совокупность нескольких однородных элементов, которое может рассматриваться как самостоятельная единица;

СУБД - Система управления базами данных, совокупность программных средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных;

ПО – программное обеспечение;

Apache – свободный кроссплатформенный HTTP-сервер;

FreeBSD – свободная Unix-подобная операционная система;

SPSS - Statistical Package for the Social Sciences (англ. «статистический пакет для социальных наук») — компьютерная программа для статистической обработки данных;

MySQL - свободная реляционная система управления базами данных;

Java - объектно-ориентированный язык программирования для написания клиентских приложений и серверного программного обеспечения;

XML - eXtensible Markup Language (англ. «расширяемый язык разметки»), язык с простым синтаксисом, удобный для создания и обработки

документов программами и удобный для чтения и создания документов человеком.

1 ОБЗОР ЛИТЕРАТУРЫ

Область применения разрабатываемого программного продукта достаточно обширна. Поэтому по данной теме уже достаточно много исследований. Например в статье Мартышенко С.Н., Мартышенко Н.С., Кустова Д.А., рассматриваются существующие универсальные программные средства для анализа данных, такие как EXCEL, SPSS, STATISTICA, STADIA и другие.

В тоже время, в статье В.Г. Андрееенкова, А.И. Орлова и Ю.Н. Толстого, которая была написана намного раньше появления данных пакетов, предпринимается попытка проанализировать процесс применения математических методов в социологических исследованиях.

В работе Николаева Я.Н. и Абрикова В.С. описаны качественные и количественные методы анализа данных, изучение которых дает ясное понимание каждого из них.

Данные источники позволили получить более полное представление об уже существующих решениях в данной теме.

2 ПОСТАНОВКА ЗАДАЧИ

Разработка сетевого ресурса для сбора и обработки социологических данных.

Областью применения разрабатываемого сайта в данном случае являются обработка показателей участников онлайн-тестирования, в том числе анонимных, накопленные в базе данных портала. Предполагается использование данного модуля для расширения возможностей сотрудников психологической службы при анализе результатов тестирования и составлении прогнозов.

Для достижения поставленной цели, необходимо следующие задачи:

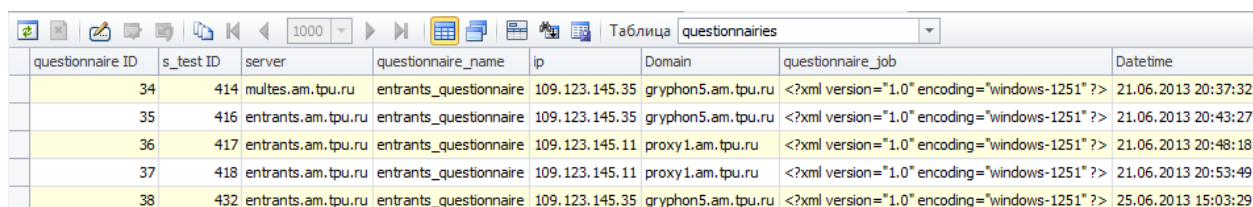
- Определить функциональность интернет ресурса, а именно какие задачи он будет решать;
- Необходимо, что бы на сайте была возможна регистрация пользователей, тестирование пользователей, а так же хранение и обработка полученной информации;
- Необходимо разработать структуру сайта и схему связности модулей;
- Разработать базу данных;
- Выбрать методы для анализа данных;

2.1 Характеристика экспериментальных данных

В качестве исходной информации имеем сведения о результатах тестирования, включающих в себя:

- 1) Анкетные данные (в случае, если тестирование не было анонимным);
- 2) Информация о соединении (IP-адрес, регион, использовавшийся браузер и т.д.);
- 3) Ответы на тестовые вопросы;
- 4) Результаты тестирования (предложенные профессии и степень соответствия им тестируемого);

Кроме того в таблицах базы присутствует информация о структуре самих тестов и ряд дополнительных сведений. Хранение всей перечисленной информации осуществляется с использованием СУБД MySQL, при этом для записи разнородных данных применяется язык разметки XML. Фрагмент одной из таблиц с записями о тестировании представлен на рисунке 1.



questionnaire ID	s_test ID	server	questionnaire_name	ip	Domain	questionnaire_job	Datetime
34	414	multes.am.tpu.ru	entrants_questionnaire	109.123.145.35	gryphon5.am.tpu.ru	<?xml version="1.0" encoding="windows-1251" ?>	21.06.2013 20:37:32
35	416	entrants.am.tpu.ru	entrants_questionnaire	109.123.145.35	gryphon5.am.tpu.ru	<?xml version="1.0" encoding="windows-1251" ?>	21.06.2013 20:43:27
36	417	entrants.am.tpu.ru	entrants_questionnaire	109.123.145.11	proxy1.am.tpu.ru	<?xml version="1.0" encoding="windows-1251" ?>	21.06.2013 20:48:18
37	418	entrants.am.tpu.ru	entrants_questionnaire	109.123.145.11	proxy1.am.tpu.ru	<?xml version="1.0" encoding="windows-1251" ?>	21.06.2013 20:53:49
38	432	entrants.am.tpu.ru	entrants_questionnaire	109.123.145.35	gryphon5.am.tpu.ru	<?xml version="1.0" encoding="windows-1251" ?>	25.06.2013 15:03:29

Рис. 1. Фрагмент таблицы

Вывод по разделу 2. Определена цель и задачи, указана их значимость, очерчены границы применения разрабатываемой подсистемы и охарактеризованы экспериментальные данные.

3. СТРУКТУРА САЙТА

3.1 Технические требования к подсистеме

Для хранения информации сайта применяется СУБД MySQL. Гибкость данной СУБД обусловлена поддержкой большого количества разных типов таблиц и поддерживаемых форматов данных. Исходя из этого разрабатываемый сайт должен взаимодействовать с СУБД MySQL, производя выгрузку данных из неё для обработки и сохраняя в базу данных ее результаты.

Для описания тестов и методик используется универсальный формат представления данных XML, что позволит, например, для создания новых тестов включаемых в портал, применять различные сторонние программы, а так же обычные редакторы текста. В том же формате происходит и хранение в базе данных информации о результатах тестирований, [1] из чего следует, что сайт, обрабатывающий результаты, должен быть приспособлен к работе с XML-файлами.

Разрабатываемый сайт должен быть размещён на HTTP-сервере apache, на платформе FreeBSD и реализовываться с использованием php, html и AJAX.

Так же сайт должен быть доступен через сеть Internet, что означает необходимость его тестирования на всех наиболее распространенных браузерах (Internet Explorer v7 и выше, Mozilla v3 и выше Opera v9 и выше и т.д.).

Сайт должен иметь разделение пользователей на различные группы, у которых будут разные функции и производить разделение прав доступа разных групп, а так же гарантировать безопасность хранения информации и защиту от несанкционированного доступа. В целях более проработки описанных выше схем взаимодействия пользователей с разрабатываемым сайтом составим сценарии взаимодействия пользователей и системы в виде вариантов использования. Варианты использования это - описание

последовательности действий, которые может осуществлять система в ответ на внешние воздействия пользователей или других программных систем [2].

3.2 Модели поведения пользователя

Работа данного сайта предполагает обработку множества данных, включая информацию, которая не предназначена для общего пользования, поэтому должно быть разграничение прав доступа и выделения групп пользователей у каждой из которых будут свои права и полномочия.

Пользователь входящий в ту или иную группу получает доступ к различным функциям сайта и компонентам базы данных в рамках квалификации пользователя.

Всех посетителей сайта можно отнести к одной из двух групп:

- 1) Администраторы;
- 2) Незарегистрированные пользователи и «гости»;

Группа администраторов должна поддерживать базу данных, заниматься резервным копированием и восстановлением информации в случае непредвиденных ситуаций. Кроме того у пользователей данной группы есть еще одна обязанность, по обеспечению безопасности, назначению и удалению прав доступа.

Однако, помимо администраторов, занятых в обеспечении функционирования тестовой системы, на сайте должен быть ограниченный набор функций для незарегистрированных пользователей, «гостей». Кроме у незарегистрированного пользователя должна быть возможность зарегистрироваться, а так же связаться с администрацией.

В целях более точного выполнения приведенных выше схем взаимодействия пользователей с сайтом составим возможные сценарии взаимодействия пользователей и сайта.

3.2.1 Сценарии поведения администратора

У администраторов, как было описано выше, есть определенные обязанности, такие как: установления прав доступа новых участников проекта, их изменение или лишение прав и доступа тех пользователей,

которые покидают проект. Кроме того в работу администраторов так же входит управление базой данных, импорт в файлы, экспорт данных из файлов при переносе или изменении базы, резервное копирование данных и восстановление базы после непредвиденных сбоев.

Рассмотрим ниже сценарии работы администратора.

Сценарий 1. Регистрация нового пользователя-администратора

Администратор осуществляет проверку и регистрацию нового пользователя-администратора. В данном сценарии предполагается, что администратор авторизован и рассматривает полученную заявку от пользователя, проверяя ее на наличие ошибок. Заявка должна быть корректно заполнена и подана на регистрацию.

Администратор обязан ознакомиться с регистрационными данными пользователя, а так же с причиной по которой пользователю нужно предоставить доступ к данным и функционалу сайта и принять решение.

Если было решено дать доступ пользователю, администратору необходимо с использованием базы данных внести пользователя в группу «Администраторы» и выдать ему уникальный логин и пароль.

Альтернативный сценарий:

Администратор, рассмотрев заявку не нашел достаточных оснований, для выдачи доступа данному пользователю. В данном случае появляется еще один сценарий.

Сценарий 2. Отказ в регистрации нового пользователя

Администратор отказывает в доступе новому пользователю, так как не нашел достаточно оснований, для выдачи прав. В данном случае администратор обязан сообщить пользователю об отказе, по указанному в заявке e-mail адресу.

Сценарий 3. Администратору необходимо сохранить базу данных

Администратор необходимо сохранить базу данных или отдельных таблиц на жесткий диск. В данном случае администратору необходимо авторизоваться под своим уникальным логином и паролем. Далее

необходимо вызвать меню сохранения компонентов базы данных, выбрать все или определенные из всех таблиц, предложенных базой данных и указать желаемый адрес сохранения файлов. Система создает в указанном каталоге XML-файлы, генерирует названия, исходя из тех названий, что были у основных таблиц и даты создания данных копий. Далее система сохраняет данные о содержимом и структуре исходных таблиц в созданные файлы в формате XML.

После всех этих операций система создает запись в логах о том, что данные были скопированы, после этого администратору выдается сообщение об успешном копировании.

Альтернативные сценарий:

По какой-то причине копирование не состоялось, либо было невозможным.

Сценарий 4. Невозможность копирования данных

Система сообщает администратору о невозможности копирования данных. Существует несколько наиболее распространенных причин, по которым такое событие могло произойти:

- Утрата связи с базой данных
- Сбой сайта
- Отсутствие необходимого количества места на жестком диске

В любом из этих случаев администратору будет выведено сообщение с указанием причин из-за которых копирование невозможно. Администратору необходимо оценить возможности устранения причин сбоя и если устранение прошло успешно, повторить процедуру копирования.

Сценарий 5. Загрузить базу данных из файла

Администратор производит перенос данных из XML-файла в базу. В этом случае администратор должен быть авторизован под своими уникальными логином и паролем. Необходимо вызвать меню записи данных, указать адрес размещения файлов. После этого система читает предложенные

в каталоге XML-файлы и выводит их список администратору, выбираются те таблицы из списка, которые должны быть внесены в базу, система создает копии выбранных таблиц и добавляет данные из файлов в базу. После чего система создает запись в логах о факте добавления данных и выводит администратору сообщение об успешном добавлении.

Альтернативные сценарий:

По какой-то причине данные сохранить было невозможно.

Сценарий 6. Невозможность сохранения данных

Система сообщает администратору о том, что сохранение данных невозможно. Такое может произойти по ряду определенных причин, наиболее частые:

- Утрата связи с базой данных
- Сбой сайта

Администратор должен произвести поиск причин сбоя и оценить возможности их устранения. Если ошибки были устранены, производится повторная процедура переноса данных.

3.2.2 Сценарии поведения пользователя «Гость»

Пользователь «Гость» получает доступ к новостям портала, внешним ссылкам, имеет возможность авторизоваться на сайте либо отправить запрос на регистрацию или сообщение администратору.

Сценарий 1. Зарегистрироваться в системе

Пользователь «Гость» отправляет запрос о регистрации новой учётной записи. Для этого гостю необходимо перейти к регистрации и заполнить регистрационную форму, система проверяет заявку по основным требованиям, после чего вносит регистрационные данные новой учетной записи в базу данных.

Альтернативный сценарий: Отказ в регистрации.

Сценарий 2. Отказ в регистрации

Такой сценарий возможен если пользователь, заполняя регистрационную форму допустил ряд ошибок, либо на данный e-mail адрес уже был зарегистрирован пользователь.

Сценарий 2. Пройти тестирование

Пользователь может выбрав любую из предложенных анкет, пройти тестирование. После прохождения теста по нажатию кнопки «Отправить» результаты тестирования будут отправлены в базу данных.

Сценарий 3. Оставить отзыв

На сайте пользователь может оставить отзыв, содержание которого будет так же отправляться на хранение и рассмотрение в базу данных.

Вывод по разделу 3. Определены роли пользователей сайта и распределены их права доступа, а так же обязанности в системе. Разработаны варианты использования компонентов сайта различными пользователями.

4 МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Статистические данные – это множество объектов, определенных признаков и случаев, которые являются их характеризующими. В нашем случае, объект исследования – множество тестируемых, а их признаки – множество показателей, которые их характеризуют: их личные данные и результаты.

Переменные – это величины, при измерение которых могут получаться различные значения.

Зависимые переменные – это такие величины, которые в процессе эксперимента нельзя поменять, их можно только измерить, а независимые переменные наоборот, являются теми величинами, которые можно поменять.

Так же существуют различные шкалы, в которых измеряются переменные. Информативность определяет различие шкал. Иерархия шкал, расположенная по возрастанию информативности: номинальная, порядковая, интервальная, шкала отношений, абсолютная. Каждая из этих шкал имеет отличие от другой, которое заключается в том, что над каждой из них можно проводить только определенное количество математических действий.

Шкала у которой вообще нет никаких арифметических действий - номинальная, а шкала у которой наоборот, самое большое количество арифметических действий - абсолютная. Проводя измерения в номинальной можно только определить к какому классу принадлежат наблюдения (объекты), а так же посчитать сколько объектов в классе – частоту и относительную частоту.

Если проводить измерения в порядковой шкале, можно определить класс принадлежности, а так же привести в порядок наблюдения. Однако эта шкала не определяет дистанцию между классами, а только то, какое из двух наблюдений предпочтительнее. Поэтому порядковые экспериментальные

данные, даже если они изображены цифрами, нельзя выполнять над ними арифметические операции, и рассматривать как числа.

У порядковых и номинальных переменных название, их называют категориальными, или группирующими, потому что с их помощью возможно разделить исследуемые объекты на подгруппы. При проведении в интервальной шкале измерений упорядочивая наблюдения возможно сделать это настолько точно, что расстояния между любыми двумя из них станут известны. Шкала интервалов единственна с точностью до линейных преобразований ($y = ax + b$), это значит, что у шкалы есть произвольная точка отсчета- условный нуль. Некоторыми из примеров таких переменных, измеренных в интервальной шкале являются: высота местности над уровнем моря, время, температура. Операцией, которую можно выполнять над переменными в данной шкале называют «определение расстояния между наблюдениями».

Шкала с точностью до преобразования вида $y = ax$ схожа с интервальной шкалой, ее называют шкалой отношений, но она единственна. Это означает, что у шкалы имеется фиксированная точка отсчета – абсолютный нуль, но произвольный масштаб измерения. Над измерениями в данной шкале можно производить любые арифметические действия, так как они являются числами.

Абсолютный нуль, и абсолютную единицу измерения (масштаб) имеет только абсолютная шкала. Числовая прямая является примером абсолютной шкалы. Данная шкала является безразмерной, поэтому измерения в ней можно использовать в качестве основания логарифма или показателя степени.

4.1 Выбор методов обработки данных

Множество статистических методов принадлежат к методам параметрической статистики, на которые распространяются предположения,

что случайный вектор переменных образует некоторое многомерное распределение, в большинстве случаев, преобразуется к нормальному распределению или сразу к нормальному распределению. Если этого не происходит необходимо использовать непараметрические методы математической статистики [3].

При решении данной задачи применение многих традиционных статистических методов нерационально и не позволяет добиться желаемого результата. Более эффективным представляется использование методов глубинного анализа данных, Data Mining. Data Mining — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [4].

Особенностью Data Mining является, как уже было отмечено, нетривиальность разыскиваемых шаблонов. Искомые шаблоны отражают незаметные, неожиданные регулярности в данных, которые составляют так называемые скрытые (глубинные) знания. Выделяются пять видов закономерностей, позволяющие выявить методы Data Mining: ассоциация, классификация, последовательность, прогнозирование и кластеризация [5].

Ассоциацию присутствует только тогда, когда несколько событий связаны друг с другом. Так же говорят о последовательности, она присутствует если несколько событий связаны друг с другом во времени событий. С помощью классификации выделяются признаки, которые характеризуют группу, к которой принадлежит рассматриваемый объект. Это производится с помощью анализа тех объектов, которые уже классифицированы и формулирования определенного набора правил. Классификация отличается от кластеризации тем, что изначально сами группы не заданы. Используя кластеризацию средства Data Mining самостоятельно выделяют разные однородные группы данных. Как правило, на основе кластеризации формируются новые продукты [6].

Историческая информация, хранящаяся в БД в виде временных рядов служит основой для всевозможных систем прогнозирования. Если построить шаблоны представляется возможным и они будут адекватно отражать динамику поведения целевых показателей, есть вероятность, что с их помощью возможно предсказывать и поведение системы в будущем. В рассматриваемом приложении на первый план выходят задачи кластеризации или иначе говоря кластерного анализа данных.

4.2 Методы Data Mining

Все методы Data Mining делятся на две группы по принципу того, как они работают с исходными данными. После Data Mining данные могут сохраняться, либо дистиллироваться для дальнейшего использования- это определяет верхний уровень в данной классификации.

В случае использования исходных данных они сохраняются в явном детализированном виде и используются на стадиях прогностического моделирования или анализа исключений. Недостаток этой группы методов в сложности анализа больших баз данных. В данную группу входят методы кластерного анализа, метод ближайшего соседа, рассуждение по аналогии.

Методы с использованием формализованных закономерностей (дистилляция шаблонов). При данной технологии один образец (шаблон) информации изымается из начальных данных и подвергается преобразованию в формальные конструкции, в этом случае вид зависит от применяемого метода Data Mining. Данный процесс проводится на ступени свободного поиска (у первой группы методов этой ступени нет). На дальнейших стадиях используются результаты стадии свободного поиска, которые значительно компактнее исходных баз данных. Кроме того конструкции этих моделей могут быть как трактуемыми аналитиком, так и "черными ящиками" (неразборчивыми). Методы этой группы: методы кросс-табуляции, логические методы, методы основанные на уравнениях, методы визуализации.

Логические методы (методы логической индукции) содержат в себе, деревья решений, нечеткие запросы и анализы. Методы этой группы оформляют найденные закономерности в достаточно понятном виде с точки зрения пользователя, поэтому они являются наиболее интерпретируемыми. Нужно заметить, что деревья решений могут быть легко преобразованы в наборы символьных правил путем генерации одного правила по пути от корня дерева до его терминальной вершины.

Методы кросс-табуляции: доверительные сети, агенты, кросс-табличная визуализация.

Методы, основанные на уравнениях демонстрируют найденные закономерности в виде математических уравнений. Данные методы могут работать только с численными переменными, а это несколько ограничивает использование методов этой группы. Так или иначе они широко используются при решении задач прогнозирования. Основные методы данной группы используют нейронные сети.

Основа другой классификации- различный подход к обучению математических моделей. Согласно ей все многообразие методов Data Mining можно разделить на две группы: статистические и кибернетические методы. Статистические методы основаны на усреднении накопленного опыта, отраженного в данных, тогда как кибернетические методы, включают в себя множество разнородных математических подходов.

Недостаток данной классификации состоит, в том, что и статистические, и кибернетические алгоритмы опираются на сопоставление статистического опыта с результатами мониторинга текущей ситуации.

Плюсом данной классификации является ее удобство для интерпретации.

Статистические методы Data mining разделены на четыре взаимосвязанных раздела:

1. Предварительный анализ природы статистических данных
2. Выявление связей и закономерностей

3. Многомерный статистический анализ (в т.ч. кластерный анализ)
4. Динамические модели и прогноз на основе временных рядов.

Арсенал статистических методов Data Mining также разделен на четыре группы методов:

1. Дескриптивный анализ и описание исходных данных
2. Анализ связей
3. Многомерный статистический анализ
4. Анализ временных рядов.

У второго направления Data Mining существует множество подходов, объединенных использованием элементов теории искусственного интеллекта. К этой группе относят такие методы:

1. Искусственные нейронные сети
2. Эволюционное программирование и генетические алгоритмы
3. Ассоциативная память (поиск аналогов, прототипов)
4. Нечеткая логика
5. Деревья решений
6. Системы обработки экспертных знаний.

Методы Data Mining могут быть разделены по задачам Data Mining. Придерживаясь такому разделению можно выделить две группы, которые решают задачи прогнозирования (классификации и кластеризации) и задачи сегментации.

В соответствии со второй классификацией по задачам методы Data Mining могут быть использованы для получения описательных и прогнозирующих результатов. Описательные методы используются для того, чтобы найти образцы или шаблоны, описывающие данные, которые более приемлемы для восприятия с точки зрения аналитика.

Методы, которые направлены на получение описательных результатов, известны как итеративные методы кластерного анализа, в том числе: алгоритм k-медианы, k-средних, иерархические методы кластерного анализа,

методы кросс-табличной визуализации, самоорганизующиеся карты Кохонена, различные методы визуализации и другие.

Прогнозирующие методы пользуются значениями одних переменных для прогнозирования неизвестных или будущих значений целевых переменных.

Методы которые направлены на получение описательных результатов: линейная регрессия, деревья решений, нейронные сети, метод ближайшего соседа, метод опорных векторов и др.

4.2.1 Методы кластерного анализа

Кластерный анализ - это многомерная статистическая процедура, которая выполняет сбор данных, содержащих информацию о выборке объектов, и затем упорядочивает объекты в однородные группы (кластеры).

Главным назначением кластерного анализа является разбиение множества исследуемых объектов и признаков на однородные группы (кластеры). Это позволяет решать задачу классификации данных и структурировать их. Методы кластерного анализа применимы даже случаях, когда речь идет о простой группировке, с образованием групп по количественному сходству. Значимое достоинство кластерного анализа состоит в том, что он позволяет осуществлять разбиение и группировку объектов по целому набору признаков.

В критерий качества классификации так или иначе входит ряд неформальных требований [7]:

- 1) взаимосвязь объектов внутри группы
- 2) объекты разных групп должны быть далеки друг от друга;
- 3) распределения объектов по группам при прочих равных условиях должны быть равномерными.

Основная цель кластерного анализа – выделить в исходных n -мерных данных такие однородные подмножества, чтобы объекты внутри групп были близки друг другу и отличались от объектов из других групп. Попытки

охарактеризовать близость объектов в многомерном пространстве признаков, сводят задачу к выделению в этом пространстве однородных групп объектов.

Для решения данной задачи было бы достаточно, если бы объекты i и j попадали в один кластер всякий раз, когда расстояние между соответствующими точками x_i и x_j было бы достаточно мало, в противном же случае они попадали в разные кластеры. В таблице 1 приведены примеры наиболее часто используемых функций расстояния.

Таблица 1. Некоторые функции расстояния, используемые в методах кластерного анализа

Название	Формула	Обозначения
1. Евклидово расстояние	$d_2(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2}$	i, j — номера объектов; p — количество переменных, характеризующих объект; k — номер признака; x_{ki} — значение k -го признака у i -го объекта; S — ковариационная матрица
2. L1-норма	$d_1(x_i, x_j) = \left[\sum_{k=1}^p x_{ki} - x_{kj} \right]$	
3. Супремум норма	$d_\infty(x_i, x_j) = \sup_{k=1,2,\dots,p} \left\{ x_{ki} - x_{kj} \right\}$	
4. Lp-норма	$d_p(x_i, x_j) = \left[\sum_{k=1}^p x_{ki} - x_{kj} \right]^{1/p}$	
5. Расстояние Махаланобиса	$d(x_i, x_j) = \sqrt{(x_i - x_j) S^{-1} (x_i - x_j)}$	
6. Расстояние Хэмминга	$d(x_i, x_j) = \sum_{k=1}^p x_{ik} - x_{jk} $	

Евклидово расстояние, геометрическое расстояние в многомерном пространстве, можно охарактеризовать как наиболее общий тип расстояния.

При кластеризации многие процедуры совершаются ступенчато. Это означает, что два, наиболее близко расположенных объекта x_i и x_j , объединяются и рассматриваются как один кластер. Это приводит к тому, что число объектов уменьшается на один (один кластер будет содержать два объекта, а оставшиеся по одному). Наиболее подходящее разбиение выбирает чаще всего сам исследователь на основе дендрограммы, которая отображает результаты группирования объектов на каждом шаге алгоритма.

Дендрограмма - n-уровневая древовидная диаграмма, каждый из уровней которой соответствует одному из шагов процесса последовательного укрупнения кластеров. Дендрограмму также называют деревом объединения кластеров, деревом иерархической структуры. Традиционно различают классификации: иерархические и иногда структурными (не иерархичные). Соответственно можно разделить алгоритмы получения этих классификаций.

Основа иерархических алгоритмов состоит в объединении в кластер наиболее близких, а затем последовательно и всё более отдалённых друг от друга элементов. Большинство из этих алгоритмов исходит из матрицы расстояний, и каждый отдельный элемент рассматривается вначале как отдельный кластер.

Общая схема такой иерархической группировки может быть представлена как повторяющееся приложение трех операций к мерам расстояния между кластерами:

- 1) найти наименьшее расстояние d_{S_1, S_2} между кластерами S_1 и S_2 ;
- 2) объединить S_1 и S_2 в один кластер, присвоив результату объединения общий индекс;
- 3) вычислить расстояние $d_{S, S_1 \cup S_2}$ от кластера $S_1 \cup S_2$ до любого другого объекта (кластера) S .

Существует ряд методов кластерного анализа, базирующихся на евклидовой мере расстояния, в том числе метод Уорда, метод ближайшего соседа, метод дальнего соседа и медианный метод.

Метод Уорда (Ward's method). В качестве целевой функции применяется сумма квадратов расстояний между каждой точкой, соответствующей объекту, и средней по кластеру, содержащему этот объект (внутригрупповая сумма квадратов отклонений). Данный метод реализуется последовательной процедурой: на каждой итерации объединяются два кластера, которые приводят к минимальному увеличению целевой функции (суммы квадратов отклонений). В результате происходит объединение

близко расположенных кластеров.

Метод ближайшего соседа. На первом шаге метода каждый объект считается отдельным кластером. Расстояние между кластерами равно расстоянию между двумя ближайшими соседями классов. На каждом последующем шаге независимо от общей формы кластера к нему присоединяются ближайшие к границе объекты.

Метод дальнего соседа. Аналогично всем вышеназванным методам на первом шаге каждый объект считается отдельным кластером, и межкластерное расстояние равно расстоянию между самыми дальними объектами. В остальном процесс повторяет метод ближайшего соседа.

Медианный метод. В данном методе расстояние между кластерами равно расстоянию между точками с медианными значениями признаков в классах.

Помимо иерархических кластерных процедур, существуют неиерархические, также называемые структурными. В них реализуется образование кластеров по принципу выделения т.н. «сгущений» - мест наибольшей концентрации точек в пространстве. К данной категории относится метод k-средних.

Алгоритмы кластерного анализа отличаются большим разнообразием. На практике в результате испытаний различных алгоритмов исследователями был сформирован ряд рекомендаций по их применению:

- 1) Различные алгоритмы кластерного анализа зачастую дают разные разбиения. Близкие результаты наблюдаются у методов группового среднего и методов центроидного разделения, а также у вариаций одного и того же метода (метод k-средних в варианте Болла и Холла варианте Мак-Кина);
- 2) Результаты классификации находятся в обратной зависимости от размерности пространства и прямой от целевого количества кластеров
- 3) Наиболее устойчивы к зашумлению данных алгоритмов метод k-средних и метод Уорда, наименее – медианный метод и метод ближнего соседа;

4) При малом числе признаков и большом числе классов лучше всего себя показали метод Уорда, метод группового среднего и метод дальнего соседа; плохо – метод k-средних и метод ближайшего соседа.

5) При большом числе признаков и малом числе классов наиболее эффективны метод Уорда и метод k-средних, наименее – медианный, центроидный и метод ближайшего соседа.

6) В широком диапазоне условий случаях достаточно хорошо работают метод дальнего соседа и метод Уорда, плохо – медианный и метод ближайшего соседа;

С учетом чувствительности к зашумлению и способности восстановлению структуры данных наилучшим является алгоритм Уорда, наихудшим – метод ближайшего соседа [8].

Кластерный анализ результатов тестирования

Методы кластерного анализа и ранее имели применение в проводимых с помощью сайта исследованиях, однако это было сопряжено с рядом неудобств, вызванных применением разных, нередко слабо согласованных программных продуктов. Отсюда следует дополнительная задача для разрабатываемого программного комплекса: осуществление проверки и верификации уже полученных результатов обработки.

Примером таких результатов можно назвать данные, полученные в работе по кластеризации данных, полученных в ходе тестирования потенциальных абитуриентов, использовавшиеся впоследствии в работе по исследованию гендерных особенностей студентов и абитуриентов различных специальностей [9]. Так, результаты кластеризации по методике Уорда психологических показателей теста FPI (Фрайбургский личностный опросник) показаны в таблице 2.

Таблица 2. Центроиды кластеров теста FPI

Показатели теста FPI	Кластер 1	Кластер 2	Кластер3

Невротичность	6,54	6,88	3,89
Спонтанная агрессивность	6,50	6,81	4,46
Депрессивность	6,11	7,25	4,04
Раздражительность	6,57	7,44	5,11
Общительность	3,39	4,19	5,43
Уравновешенность	6,07	3,72	6,46
Реактивная агрессивность	7,07	6,59	5,50
Застенчивость	6,46	7,28	4,96
Открытость	6,21	8,19	6,46
Экстраверсия – интроверсия	4,82	5,09	6,07
Эмоциональная лабильность - стабильность	6,14	7,53	4,32

Из полученных данных был сделан вывод, что в первый кластер попали более замкнутые тестируемые с низкой общительностью и с выраженной реактивной агрессивностью, ко второму - неуравновешенные, не уверенные в себе, депрессивные, раздражительные, чрезмерно щепетильные, но в тоже время доверительно-откровенно взаимодействующие с окружающими людьми при высоком уровне самокритичности. Третий кластер составили психологически и эмоционально устойчивые люди, для которых характерны неимпульсивность, обдуманная реакция, уверенность, без внутренних противоречий и сомнений.

Вывод по разделу 4. Рассмотрены способы решения поставленной задачи и выбраны методы, наиболее адекватные её специфике. Перечислены методы анализа данных, желательные для программной реализации. Приведён пример применения выбранных методов к решению психологических задач.

5 ПРОГРАММНЫЕ СРЕДСТВА

На данный момент развитие программного обеспечения развито на высоком уровне, что значительно облегчает «ручную» реализацию выбранных методик обработки данных.

В настоящее время в социальных исследованиях широко используются различные математико-статистические методы обработки данных, реализованные в программных приложениях с применением современных информационных технологий.

Компьютерная обработка данных предполагает некоторое математическое преобразование данных с помощью определенных программных средств. Для этого необходимо иметь представление как о математических методах обработки данных, так и о соответствующих программных средствах [10].

Наиболее популярными являются пакеты статистического анализа данных, такие как SPSS, Statistica, STATGRAPHICS. Однако специфика задачи накладывает свои ограничения, делающие использование данных программных продуктов нежелательным:

- 1) Предпочтение программных продуктов, созданных на веб-ориентированных языках (Java, Python и т.д.);
- 2) Ограничения, накладываемые структурой сайта;
- 3) Предпочтение программных продуктов с открытым исходным кодом и широкими возможностями доработки;
- 4) Предпочтение свободного программного обеспечения.

Всё вышесказанное обращает нас к наиболее распространённым open source продуктам в области статистического анализа и Data Mining (интеллектуальный анализ данных).

5.1 Оценка программных средств Data Mining

К числу наиболее популярных и детально разработанных приложений в данной категории можно отнести Weka, RapidMiner и KNIME (написаны на

Java), Rattle (основан на R) и Orange (интегрирован с Python). В этот перечень намеренно не были включены узкоспециализированные продукты.

Вывод по разделу 5. Выбрано программное обеспечение для решения задачи, оценены его характеристики. По завершении данного этапа разработка приложения перешла в стадию итоговой программной реализации.

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
8Б21	Ковригина Таисия Сергеевна

Институт	Институт кибернетики	Кафедра	Прикладной математики
Уровень образования	бакалавриат	Направление/специальность	Прикладная математика и информатика

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Работа с информацией, представляющей собой технические тексты, методы кластерного анализа данных, а также разработка многофункционального сайта.
2. Нормы и нормативы расходования ресурсов	
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	Планирование этапов разработки программы, определение трудоемкости.
2. Планирование и формирование бюджета научных исследований	
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	

Перечень графического материала (с точным указанием обязательных чертежей):

1. Оценка конкурентоспособности технических решений
2. График проведения и бюджет НИ
3. Оценка ресурсной, финансовой и экономической эффективности НИ

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент	Николаенко Валентин Сергеевич	-		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8Б21	Ковригина Таисия Сергеевна		

Введение

Данный раздел является обязательной частью данной работы, представляющий собой детальный анализ экономической эффективности проекта.

При организации процесса реализации конкретного проекта необходимо рационально планировать занятость каждого из его участников и сроки проведения отдельных работ. Это можно сделать при помощи линейного графика работ. Для его построения сначала определим полный перечень проводимых работ, их продолжительность и исполнителей. Полученные данные сведены в таблице 3.

Таблица 3 - Перечень работ и продолжительность их выполнения

Этапы работы	Исполнители	Загрузка исполнителя НР, %	Загрузка исполнителя И, %
Постановка целей и задач, получение исходных данных	НР	50	50
Составление и утверждение ТЗ	НР, И	100	10
Подбор и изучение материалов по тематике	НР, И	40	100
Разработка календарного плана	НР, И	100	10
Обсуждение литературы	НР, И	30	100
Выбор структурной схемы системы и используемых математических методов	НР, И	60	100
Выбор методов и программных средств реализации	НР, И	20	100
Проведение расчетов и валидация модели	И		100
Создание программного продукта	И		100
Оформление расчетно-пояснительной записки	И		100
Подведение итогов	НР, И	45	100

6.1 Продолжительность этапов работ

Расчет продолжительности этапов работ определены опытно-статистическим экспертным методом.

Определим ожидаемое время проведения работ, длительность этапов в рабочих и календарных днях, по формулам:

$$t_{ож} = \frac{3 \cdot t_{min} + 2 \cdot t_{max}}{5}$$

где t_{min} – минимальная продолжительность работы, дн.;

t_{max} – максимальная продолжительность работы, дн.;

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д}$$

где $t_{ож}$ – продолжительность работы, дн.;

$K_{ВН}$ – коэффициент выполнения работ, учитывающий влияние внешних факторов на соблюдение предварительно определенных длительностей, в частности, возможно $K_{ВН} = 1$;

$K_{Д}$ – коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ ($K_{Д} = 1-1,2$).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{КД} = T_{РД} \cdot T_{К}$$

где $T_{КД}$ – продолжительность выполнения этапа в календарных днях;

Возьмем $K_{Д} = 1,1$.

$T_{К}$ – коэффициент календарности, позволяющий перейти от длительности работ в рабочих днях к их аналогам в календарных днях, и рассчитываемый по формуле

$$T_{К6} = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}} = \frac{365}{365 - 52 - 10} = 1,205$$

где $T_{КАЛ}$ – календарные дни ($T_{КАЛ} = 365$);

$T_{ВД}$ – выходные дни ($T_{ВД} = 52$);

$T_{\text{ПД}}$ – праздничные дни ($T_{\text{ПД}} = 10$).

В таблице 2 приведены продолжительности этапов работ и их трудоемкости по исполнителям, занятым на каждом этапе.

Таблица 4 - Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.- дн.			
		$t_{\min}$	$t_{\max}$	$t_{\text{ож}}$	$T_{\text{рд}}$		$T_{\text{кд}}$	
					НР	И	НР	И
1	2	3	4	5	6	7	8	9
Постановка целей и задач, получение исходных данных	НР	2	3	2,4	1,32	1,32	1,59	1,59
Составление и утверждение ТЗ	НР, И	1	3	1,8	1,98	0,20	2,39	0,24
Подбор и изучение материалов по тематике	НР, И	15	22	17,8	7,83	19,58	9,44	23,59
Разработка календарного плана	НР, И	2	3	2,4	2,64	0,26	3,18	0,32
Обсуждение литературы	НР, И	3	6	4,2	1,39	4,62	1,67	5,57
Выбор структурной схемы системы и используемых математических методов	НР, И	8	15	10,8	7,13	11,88	8,59	14,32
Выбор методов и программных средств реализации	НР, И	5	9	6,6	1,45	7,26	1,75	8,75
Проведение расчетов и валидация модели	И	12	18	14,4	0,00	15,84	0,00	19,09
Создание программного продукта	И	15	23	18,2	0,00	20,02	0,00	24,12
Оформление расчетно-пояснительной записки	И	5	7	5,8	0,00	6,38	0,00	7,69
Подведение итогов	НР, И	3	6	4,2	2,08	4,62	2,51	5,57
Итого:				88,6	25,82	91,98	31,11	110,84

Величины трудоемкости этапов по исполнителям $T_{\text{кд}}$ позволяют построить линейный график осуществления проекта (табл. 5).

Таблица 5 - Линейный график работ



6.2. Бюджет научно-технического исследования

6.2.1. Расчет материальных затрат

Так как для написания ВКР не требовалась покупка какого-либо материального оборудования и лицензий на ПО (использовались оборудование и лицензии университета), то к данной статье расходов можно отнести только расходы на распечатку материалов (прим 300 листов за время выполнения проекта), ручки, блокноты и ТЗР, см. табл. 6.

Таблица 6 - Расчет затрат на материалы

Наименование материалов	Цена за ед., руб.	Кол-во	ед	Сумма, руб.
Распечатка листов А4	2,5	300	шт	750
Ручка	17,5	10	шт	175
тетрадь, 36 л.	20	5	шт	100
ТЗР		10	%	102,5
Итого:				1127,5

6.2.2. Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{\text{эл.об.}} = P_{\text{об}} \cdot t_{\text{об}} \cdot \text{Ц}_{\text{э}}$$

где $P_{\text{об}}$ – мощность, потребляемая оборудованием, кВт;

$\text{Ц}_{\text{э}}$ – тариф на 1 кВт·час; Для ТПУ $\text{Ц}_{\text{э}} = 5,257$ руб./кВт·час (с НДС).

$t_{\text{об}}$ – время работы оборудования, час.

$$t_{\text{об}} = T_{\text{рд}} * K_t,$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{\text{рд}}$, определяется исполнителем самостоятельно. В ряде случаев возможно определение $t_{\text{об}}$ путем прямого учета, особенно при ограниченном использовании соответствующего оборудования.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{\text{об}} = P_{\text{ном.}} * K_c$$

где $P_{\text{ном.}}$ – номинальная мощность оборудования, кВт;

$K_c \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности $K_c = 1$.

Пример расчета затраты на электроэнергию для технологических целей приведен в таблице 7.

Таблица 7 - Затраты на электроэнергию технологическую

Наименование оборудования	K_t	Время работы оборудования $t_{\text{об}}$, час	Потребляемая мощность $P_{\text{об}}$, кВт	Затраты $\text{Э}_{\text{об}}$, руб.
Персональный компьютер	0,9	662,4	0,3	1044,67
принтер	0,01	7,36	0,1	3,87
Итого:				1048,54

6.2.3. Расчет заработной платы

Данная статья расходов включает заработную плату научного руководителя и инженера (в его роли выступает исполнитель проекта), а также премии, входящие в фонд заработной платы. Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада исполнителя. Величины месячных

окладов по нормам ТПУ для научного руководителя принимается равным 33 162,87р., а для студента-исполнителя – 14 874,45р.

Среднедневная тарифная заработная плата ($ЗП_{\text{дн-т}}$) рассчитывается по формуле:

$$ЗП_{\text{дн-т}} = МО/24,83$$

учитывающей, что в году 298 рабочих дней и, следовательно, в месяце в среднем 24,83 рабочих дня при шестидневной рабочей неделе.

Расчеты полной заработной платы для обоих участников проекта, с учетом ряда коэффициентов ($K_{\text{пр}} = 1,1$; $K_{\text{доп.ЗП}} = 1,188$; $K_p = 1,3$), приведены в таблице 8.

Таблица 8 - Затраты на заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./раб.день	Затраты времени, раб.дни	Коэффициент	Фонд з/платы, руб.
НР	33 162,87	1617,70	26,00	1,70	71 453,59
И	14 874	599,05	92,00	1,70	93 627,73
Итого:					165 081,32

6.2.4. Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН) включают отчисления в пенсионный фонд, социальное и медицинское страхование, и составляют 30 % от полной заработной платы по проекту (табл. 9):

$$C_{\text{соц.}} = C_{\text{зп}} * 0,3.$$

Таблица 9 - Затраты на ЕСН

Исполнитель	ЕСН
НР	21 436,08
И	28 088,32
Итого:	49 524,40

6.2.5. Расчет амортизационных расходов

Для расчета амортизации оборудования используется формула:

$$C_{AM} = \frac{H_A * C_{OB} * t_{pf} * n}{F_D},$$

где H_A – годовая норма амортизации единицы оборудования;

C_{OB} – балансовая стоимость единицы оборудования с учетом ТЗР.

F_D – действительный годовой фонд времени работы соответствующего оборудования

t_{pf} – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

n – число задействованных однотипных единиц оборудования.

Таблица 10 - Амортизационные затраты

Наименование оборудования	год фонд врем F_D	Факт Время работы оборудования t_{pf} , час	H_A	$C_{об}$	$C_{ам}$
Персональный компьютер	2384	662,4	0,33	44000,00	4075,17
МФУ	2384	7,36	0,40	9350,00	11,55
Итого:					4086,71

6.2.6 Расчет общей себестоимости разработки

Определим общую себестоимость. Табл. 11

Таблица 11 - Смета затрат на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{мат}$	1127,5
Основная заработная плата	$C_{зп}$	165 081,32
Отчисления в социальные фонды	$C_{соц}$	49 524,40
Расходы на электроэнергию	$C_{эл.}$	1048,54
Амортизационные отчисления	$C_{ам}$	4086,71
Итого:		242 955,32

6.2.7. Расчет прибыли

Прибыль от реализации проекта принимается в размере 15 % от полной себестоимости проекта.

Прибыль 36443,30

Вывод: Основываясь на данных, полученных в пунктах 2.1 – 2.7 был рассчитан бюджет затрат научно-исследовательской работы, а так же была определена прибыль от реализации проекта.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8Б21	Ковригина Таисия Сергеевна

Институт	Институт кибернетики	Кафедра	Прикладной математики
Уровень образования	Бакалавриат	Направление/специальность	Прикладная математика и информатика

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Рабочая зона –помещение, рабочее место с персональным компьютером.
--	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Производственная безопасность 1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности: – физико-химическая природа вредности, её связь с разрабатываемой темой; – действие фактора на организм человека; – приведение допустимых норм с необходимой размерностью (со ссылкой на соответствующий нормативно-технический документ); – предлагаемые средства защиты; – (сначала коллективной защиты, затем – индивидуальные защитные средства). 1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения в следующей последовательности: – механические опасности (источники, средства защиты); – термические опасности (источники, средства защиты); – электробезопасность (в т.ч. статическое электричество, молниезащита – источники, средства защиты); – пожаровзрывобезопасность (причины, профилактические мероприятия, первичные средства пожаротушения).	1.1. Анализ выявленных вредных факторов: - Повышенный уровень электромагнитных излучений - Отклонение показателей микроклимата - Недостаточная освещенность рабочей зоны Повышенный уровень шума на рабочем месте 1.2. Анализ выявленных опасных факторов: - Статическое электричество - Короткое замыкание - Пожароопасность
2. Экологическая безопасность: – защита селитебной зоны – анализ воздействия объекта на атмосферу (выбросы); – анализ воздействия объекта на гидросферу	Анализ негативного воздействия на окружающую природную среду: утилизация компьютеров и другой оргтехники

(сбросы); – анализ воздействия объекта на литосферу (отходы); – разработать решения по обеспечению экологической безопасности со ссылками на НТД по охране окружающей среды.	
3. Безопасность в чрезвычайных ситуациях: – перечень возможных ЧС при разработке и эксплуатации проектируемого решения; – выбор наиболее типичной ЧС; – разработка превентивных мер по предупреждению ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.	Возможные чрезвычайные ситуации: - Пожар
4. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	- Рабочее место при выполнении работ сидя регулируется ГОСТом 12.2.032 – 78

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
	Антоневич Ольга Алексеевна	к.б.н		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8Б21	Ковригина Таисия Сергеевна		

Введение

Данная работа посвящена алгоритмическому и программному обеспечению информационной системы для поддержки социальных исследований. При этом доступ к предоставленным возможностям осуществляется с персональных компьютеров, компьютеров размещённых в лаборатории. Информационная система развёртывается на аппаратном обеспечении (сервере), принадлежащем научной группе, проводящей исследование, и объединяет в себе базу данных, почтовый сервис, интерфейсы и набор инструментария для пользователей с различным уровнем полномочий. При этом доступ к предоставленным возможностям осуществляется как с персональных компьютеров, так и с размещённых в лаборатории.

Вопросы безопасности, равно как и вопросы экологической безопасности и некоторых правовых аспектов организации социологического исследования рассмотрены в данной работе, с привлечением как существующих ГОСТов и правовых норм по безопасности, так и с использованием методических указаний и рекомендаций по проведению массовых социологических исследований.

7.1. Производственная безопасность

В ходе работы полевых социологических сотрудник может подвергаться воздействию ряда вредных и опасных факторов. Данные факторы перечислены в таблице 12 в соответствии с ГОСТ 12.0.003-74 [11] «Опасные и вредные производственные факторы. Классификация».

Таблица 12 - Опасные и вредные факторы при проведении социального исследования

Вид работ	Факторы (по ГОСТ 12.0.003-74)		Нормативные документы
	Вредные	Опасные	
Обработка накопленных данных	- Отклонение показателей микроклимата в помещении - отсутствие или недостаток освещённости - Повышенный уровень шума на рабочем месте	- электрический ток - статическое электричество - пожароопасность	1) Шум. Общие требования безопасности устанавливаются ГОСТ 12.1.003–83 ССБТ [12]. 2) Показатели микроклимата устанавливаются СанПиН 2.2.2.548-96 [13]. 3) Нормы освещения устанавливаются СанПиН 2.2.1/2.1.1.1278–03 [14]. 4) Допустимые уровни напряженности электростатических полей устанавливается ГОСТ 12.1.045–84 ССБТ [15]. 5) ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования [17]. 6) Электробезопасность устанавливается по ГОСТ 12.1.038–82 ССБТ [16].

7.1.1. Анализ выявленных вредных факторов при разработке и эксплуатации проектируемого решения

Отклонение показателей микроклимата

Микроклимат в помещении оказывает значительное влияние на жизнь и здоровье сотрудников, проводящих длительное время на рабочем месте. В природе влажность, температура и скорость движения воздуха являются переменными факторами, за пределами строений они нередко принимают вредные и опасные для жизни и здоровья человека значения по естественным, но внутри помещения должны поддерживаться показатели, обеспечивающие длительное безопасное нахождение внутри помещения без

средств защиты. Кроме того любые отклонения от приемлемого температурного диапазона влияют на кровеносную систему, т.к. если температура выше нормы, кровеносные сосуды расширяются, при понижении температуры кровеносные сосуды соответственно сужаются. На терморегуляцию организма влияет также влажность воздуха.

Согласно нормативам СанПиН 2.2.4.548-96 [13] по микроклимату для работ, производимых сидя, без физического напряжения, в соответствии с нормативами в летний период — от 23 до 25°C, в зимнее время температура воздуха должна составлять от 22 до 24°C. Относительная влажность воздуха на постоянном рабочем месте должна держаться в диапазоне 40-60%, нормативное значение скорости движения воздуха - 0,1 м/с.

Поддержание указанных параметров в помещении должно осуществляться посредством применения систем отопления и вентиляции, а также, при необходимости, увлажнителей воздуха, кроме того, в соответствии со сводом правил СП 60.13330.2012"СНиП 41-01-2003 [13].

Недостаточная освещённость рабочей зоны

Как недостаточная, так и неравномерная яркость рабочей зоны, избыточная освещённость, мерцание и блики являются причиной повышенного напряжения глаз и вызывают целый ряд последствий от ухудшения общего самочувствия и головных болей до необратимого падения зрения и развития заболеваний глаз. Дополнительным фактором риска является специфика работы с ЭВМ.

В силу этого по нормативам САНПИН 2.2.2/2.4.1340-03 [14] "Гигиенические требования к персональным электронно-вычислительным машинам и организации работы" освещенность в зоне размещения рабочего документа должна находиться в диапазоне 300-500 лк. Освещенность поверхности экрана не должна превышать 300 лк, кроме того освещение не должно создавать бликов. Следует ограничивать неравномерность распределения яркости в поле зрения сотрудника, притом соотношение яркости между рабочими поверхностями не должно превышать 5:1, а между

рабочими поверхностями и предметами обстановки — 10:1. Коэффициент пульсации не должен превышать 5%.

Рассмотрим офисное помещение, в котором производились работы, с размерами: длина $A = 5$ м, ширина $B = 7$ м, высота $H = 4$ м. Всего имеется шесть светильников, по 4 лампы в каждом. Фактическая освещенность рассчитывается по следующей формуле:

$$E_{\phi} = \frac{N \cdot n \cdot \Phi_{\text{см}} \cdot y}{S \cdot K \cdot z}, \quad (1)$$

где N – число светильников, шт; n – число ламп в светильнике, шт; $\Phi_{\text{см}}$ – световой поток люминесцентной лампы, Лм (при мощности 11Вт – 750лм); y – коэффициент использования светового потока (для исследуемого помещения – 0.8); S – площадь помещения, м^2 ; k – коэффициент запаса (помещения с малым выделением пыли – 1,5); z – коэффициент неравномерного освещения (для люминесцентных ламп – 1,1).

Получаем

$$E_{\phi} = \frac{6 \cdot 4 \cdot 750 \cdot 0.8}{35 \cdot 1.5 \cdot 1.1} = 249 \text{ (Лк)}.$$

Отличие от нормированного уровня

$$\Delta E = \frac{E_{\phi} - E_{\text{норм}}}{E_{\text{норм}}} \cdot 100\%,$$

$$\Delta E = \frac{249 - 300}{300} \cdot 100\% = 17\%.$$

В результате получаем $-10\% \leq 17\% \leq +20\%$. Полученное значение попадает в необходимый интервал, значит, нормы освещенности в рабочем помещении соблюдаются.

Повышенный уровень шума на рабочем месте

Шум создает влиятельную нагрузку на нервную систему человека, оказывая на него психологическое воздействие. Шумовой фон провоцирует увеличение содержания в крови гормонов стресса, таких как, норадреналин и адреналин, кортизол.

Шумовой фон помещения создают шесть одновременно работающих компьютеров. Периодически возникает шум, исходящий от принтера или телефонных аппаратов. Допустимый уровень шума для работы программиста, в соответствии с ГОСТ 12.1.003–83 ССБТ [12] представлен в таблице 13. Нормы соблюдаются.

Таблица 13 Предельно допустимые уровни звукового давления ГОСТ 12.1.003–83 ССБТ

Вид трудовой деятельности/ Частоты	Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц								
	31,5	63	125	250	500	1000	2000	4000	8000
Научная деятельность, конструирование и проектирование, программирование, Рабочие места проектно-конструкторских бюро, программистов вычислительных машин и т.д.	86	71	61	54	49	45	42	40	38

Повышенный уровень электромагнитных излучений

В данной работе источником электромагнитного излучения является персональный компьютер.

Степень и характер воздействия ЭМП на организм человека зависят: от интенсивности излучения; частоты колебаний; поверхности тела, облучаемого и т.д. [15].

Для соблюдения нормативов следует руководствоваться правилами [15]:

- выбирать монитор с жидкокристаллическим экраном;
- по возможности располагать монитор в углу помещения. Так стены будут поглощать электромагнитное излучение, испускаемые боковыми и задними стенками;

- выключать монитор, даже если отходите ненадолго;
- монитор должен стоять на расстоянии вытянутой руки от кресла;
- по возможности системный блок расположить как можно дальше от вас;
- выключать компьютер, если больше не собираетесь им пользоваться;
- по возможности сокращать время, проводимое за компьютером.

Деятельность программиста проходит перед монитором, поэтому необходимо чаще делать перерыв. Для этого можно просто пройтись. Помимо опасности от электромагнитных волн излучение от монитора может нести опасность возникновения ряда глазных заболеваний, таких как близорукость или сухость глаз [15]. Согласно СанПиНу 2.2.2/2.4.1340-03 [15] приемлемые временные допустимые уровни ЭМП, создаваемых ПЭВМ на рабочих местах, указаны в таблице 14.

Таблица 14. Временные допустимые уровни ЭМП, создаваемых ПЭВМ по СанПиНу 2.2.2/2.4.1340-03

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	в диапазоне частот 5 Гц - 2 кГц	25 В/м
	в диапазоне частот 2 кГц - 400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц - 2 кГц	250 нТл
	в диапазоне частот 2 кГц - 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500

7.1.2. Анализ выявленных опасных факторов при разработке и эксплуатации проектируемого решения

Риск поражения электрическим током при эксплуатации ПЭВМ

В связи с тем, что для работы персональных ЭВМ и периферийных устройств используется электрическая энергия, их эксплуатация должна соответствовать «Правилам технической эксплуатации электроустановок потребителей» и «Правилам техники безопасности при эксплуатации электроустановок потребителей». ЭВМ являются потенциальными источниками опасности поражения человека электрическим током, в первую очередь, при неправильном подключении ЭВМ к сети.

С целью обеспечения безопасной и надежной работы, снижения вероятности возникновения аварийных ситуаций необходимо обеспечить выполнение ряда условий. В соответствии с СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» пространство для размещения рабочих мест с ЭВМ, должно иметь защитное заземление (зануление) в соответствии с техническими требованиями по эксплуатации электроустановок и вычислительной техники. Рабочие места с ЭВМ не должны размещаться поблизости от силовых кабелей, трансформаторов, технологического оборудования, создающего помехи в работе ЭВМ. Поскольку непосредственно на ПЭВМ должно подаваться стабилизированное электропитание (с отклонением от 220 В не более —10 % +15 %), подачу электроэнергии в компьютерные помещения следует осуществлять от отдельного независимого источника питания. Согласно требованиям ПУЭ корпуса всех распределительных (вводных) щитов в зданиях, сооружениях, квартирах должны быть занулены и заземлены; на вводе в здание должна выполняться система уравнивания потенциалов.

Повышенный уровень статического электричества

Опасность возникновения статического электричества проявляется в возможности образования электрической искры и вредном воздействии его на человеческий организм, и не только в случае непосредственного контакта с зарядом, но и за счет действий электрического поля, которое возникает при заряде. При включенном питании компьютера на экране дисплея накапливается статическое электричество. Электрический ток искрового разряда статического электричества мал и не может вызвать поражение человека. Тем не менее, вблизи экрана электризуется пыль и оседает на нем. В результате чего искажается резкость восприятия информации на экране. Кроме того, пыль попадает на лицо работающего и в его дыхательные пути.

Основные способы защиты от статического электричества следующие: заземление оборудования, увлажнение окружающего воздуха. Также целесообразно применение полов из антистатического материала.

7.2. Экологическая безопасность

По закону утилизация компьютерной техники в России в настоящее время является обязательной процедурой для организаций, а в ряде зарубежных стран и для физических лиц. Производиться утилизация должна в соответствии с федеральными законами от 24.06.1998 г. № 89-ФЗ «Об отходах производства и потребления» и от 26.03.1998 № 41-ФЗ «О драгоценных металлах и драгоценных камнях», а также целым рядом подзаконных актов, постановлений и инструкций.

Списанная офисная техника, в соответствии с формулировкой закона также является отходом третьего класса опасности (т.е. умеренно опасным). Соответственно, для утилизации её можно передавать лишь компаниям, имеющим необходимую лицензию в соответствии со ст. 17 Федерального закона «О лицензировании отдельных видов деятельности» № 128-ФЗ от 08.08.2001 г. При этом, согласно статье 8.2 КоАП РФ, «несоблюдение экологических и санитарно-эпидемиологических требований при сборе, накоплении, использовании, обезвреживании, транспортировании, размещении и ином обращении с отходами производства и потребления» влечет наложение штрафа.

7.3. Безопасность в чрезвычайных ситуациях

Чрезвычайными ситуациями в подобных помещениях могут быть пожары. Основы пожарной безопасности определены по ГОСТ 12.1.004-91[9] и ГОСТ 12.1.010-76[18].

По пожарной и взрывопожарной опасности в помещениях производственного и складского назначения независимо от их функционального назначения, помещение, в котором производились работы, относится к категории Д – пониженная пожароопасность [18].

Основные причины возникновения пожаров:

- перегрузка электросети;
- поджог;
- курение в неположенных местах.

7.3.1. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС

Пожарная безопасность обеспечивается системой пожарной защиты и системой предотвращения пожара.

Мероприятия по предотвращению пожара [18]:

- содержать помещение в чистоте, убирать своевременно мусор. По окончании работы проводить влажную уборку всех помещений;
- работа должна проводиться только при исправном электрооборудовании;
- на видном месте должен быть вывешен план эвакуации из корпуса с указанием оборудования, которое нужно эвакуировать в первую очередь;
- уходящий из помещения последним должен проверить выключены ли нагревательные приборы, электроприборы и т.д. и отключение силовой и осветительной электрической сети.

Соблюдение организационных мероприятий [18]:

- правильная эксплуатация приборов, установок;
- правильное содержание помещения;
- противопожарный инструктаж сотрудников аудитории;
- издание приказов по вопросам усиления ПБ;
- организация добровольных пожарных дружин, пожарно-технических комиссий;
- наличие наглядных пособий и т.п.

Действия в случае возникновения пожара [18]:

- при наличии телефона, «112» или «01» и сообщить о пожаре и своем местоположении;

- не входить в места с высокой концентрацией дыма и видимостью менее, чем 10 метров.

Если имеется возможность выйти из помещения (здания) наружу [18]:

- покинуть помещение, используя запасные и основные пути эвакуации;
- попутно отключить электроэнергию;
- передвигаться к выходу на четвереньках, при этом закрывая рот и нос подручными средствами защиты;
- плотно закрыть дверь при выходе.

Если дым и пламя в соседних помещениях не позволяет выйти наружу [18]:

- не поддаваться панике;
- проверить возможности спуститься по пожарной лестнице или выйти на крышу;
- при отсутствии возможности эвакуироваться для защиты от дыма и тепла необходимо загерметизировать своё помещение:
 - закрыть плотно двери, заткнуть щели тканью, желательно мокрой;
 - закрыть окна и форточки.
- при наличии воды, постоянно смачивать двери и пол.
- при задымлении помещения, передвигаться только на четвереньках, прикрыв рот и нос влажным носовым платком или рукавом, в сторону окна и находиться возле окна, при этом привлекать к себе внимание людей на улице.

Пожарная безопасность в ТПУ обеспечивается в соответствии с требованиями ФЗ пожарной безопасности № 69-ФЗ от 21.12.1994 г., правилами противопожарного режима в РФ.

Рабочее место располагается в главном корпусе ТПУ 203 аудитория. На рисунке 1 представлен план эвакуации второго этажа главного корпуса, а также условные обозначения, используемые в схеме эвакуации.

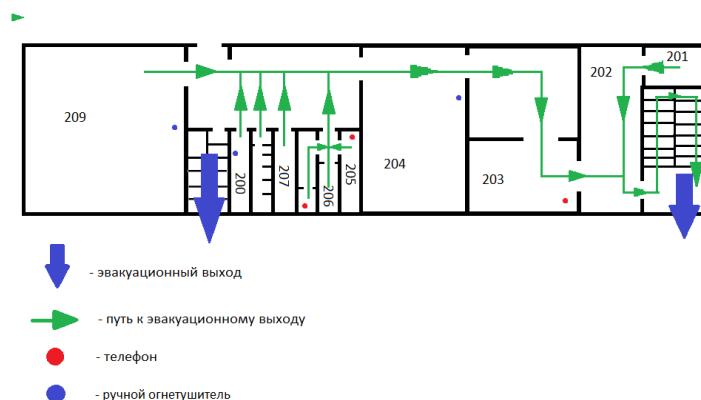


Рис.2. План эвакуации ГК ТПУ

7.4. Правовые и организационные вопросы обеспечения безопасности

Площадь одного рабочего места с компьютером должна быть не менее 6 м². При размещении рабочих мест с персональными компьютерами должны учитываться расстояния между рабочими столами с мониторами [19].

Помещения с компьютерами обязательно должны быть оборудованы системами эффективной приточно-вытяжной вентиляции отопления и кондиционирования воздуха [19].

Внутренняя отделка интерьера помещений с компьютерами должна быть сделана при использовании диффузно-отражающих материалов с коэффициентами отражения для потолка от 0,7 до 0,8; для стен от 0,5 до 0,6; для пола от 0,3 до 0,5. В помещениях с эксплуатацией компьютеров поверхность пола должна быть нескользкой, ровной и удобной для влажной уборки, а также иметь антистатические свойства [19].

В обязательном порядке в помещении должны находиться углекислотный огнетушитель для тушения пожара и аптечка первой медицинской помощи.

Согласно СанПиН 2.2.2.548-96 [20] при 8-ми часовой рабочей смене на ВДТ и ПЭВМ перерывы в работе должны составлять от 10 до 20 минут каждые два часа работы [20]. В перерывах, согласно СанПиН

2.2.2/2.4.1340-03 [19], рекомендуется проводить комплекс упражнений для глаз.

ЗАКЛЮЧЕНИЕ

В результате выполнения данной работы изучены методы анализа многомерных данных, методы data mining, в том числе методы кластерного анализа. Изучены методы, которые могут быть использованы для решения поставленной задачи. Разработаны модели поведения потенциальных пользователей. На основании этого были подобраны подходящие методы компьютерной реализации и набор программного обеспечения для данного проекта. Рассмотрены открытые программные продукты для data mining.

СПИСОК ПУБЛИКАЦИЙ СТУДЕНТА

1. Ковригина Т.С. Модель информационной системы для поддержки социальных исследований// Технологии Microsoft в теории и практике программирования: сборник трудов XIII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых, Томск, 2016. - Томск: ТПУ, 2016

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. 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 г. – Т.2 - С. 12-14

2. Computer-Aided Software Engineering Club, Варианты использования (Use Case) [Электронный ресурс]

URL: http://www.caseclub.ru/articles/use_case.html доступ свободный

3. Лаборатория статистических исследований Кубанского государственного университета, Обзор методов статистического анализа данных [Электронный ресурс] URL: <http://www.statlab.kubsu.ru/node/4>

4. Григорий Пятецкий-Шапиро, Data Mining и перегрузка информацией // Вступительная статья к книге: Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. 3-е изд. перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с. С.13.

5. Типы закономерностей, выявляемых методами Data Mining [Электронный ресурс]

URL: http://fsecrets.ru/2010/10/типы_закономерностей_выявляемых_мет/ доступ свободный

6. Data Mining: Concepts, Models, Methods and Algorithms/ Mehmed Kantardzic – New Jersey, 2011.– С. 249-253.

7. Мандель И. Д., Черный Л. М. Экспериментальное сравнение алгоритмов кластер-анализа // Автоматика и телемеханика, 1988, № 11. – 252 с.

8. Берестнева О.Г., Е.А. Муратова Компьютерный анализ данных. – Томск: ТПУ, 2010. – 144 с.

9. Берестнева О. Г. , Шпак Ю. К. , Ясюкевич Ю. В. Анализ ценностных ориентаций и психологических особенностей студентов экономических [Электронный ресурс] // Интернет журнал Науковедение. - 2013 - №. 6 (19). - Режим доступа: <http://naukovedenie.ru/PDF/33PVN613.pdf>

10. Мандель И. Д., Черный Л. М. Экспериментальное сравнение алгоритмов кластер-анализа // Автоматика и телемеханика, 1988, № 11. – 252 с.

11. ГОСТ 12.0.003-74. ССБТ. Опасные и вредные производственные факторы. Классификация.

12. ГОСТ 12.1.003–83 ССБТ. Шум. Общие требования безопасности.

13. СанПиН 2.2.2.548-96. Гигиенические требования к микроклимату производственных помещений.

14. СанПиН 2.2.1/2.1.1.1278–03. Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий.

15. ГОСТ 12.1.045–84 ССБТ. Электростатические поля. Допустимые уровни на рабочих местах и требования к проведению контроля.

16. ГОСТ 12.1.038–82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.

17. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.

18. ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования.

19. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».

20. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.

21. Приложения

Приложение А1 Главная страница сайта

```
<head>

<title>Социологическое on-line анкетирование и психологическое тестирование</title>

<meta charset="utf-8">

<link type="text/css" rel="stylesheet" href="styles/style.css">


</head>

<body>

<div id="wrap">

  <div id="header">

    <h1><a href="index.html">Социологическое on-line анкетирование и психологическое
тестирование</a></h1>


    <ul id="nav">

      <li><a href="index.html">Главная</a><span>/</span></li>

      <li><a href="anceti.html">Анкеты</a><span>/</span></li>

      <li><a href="avtr.html">Авторизация</a><span>/</span></li>

      <li><a href="reg.html">Регистрация</a><span>/</span></li>

      <li><a href="contact.html">Обратная связь</a></li>

    </ul>

  </div>

  <div id="intro">

    <h3>Важную роль в обеспечении благополучия человека и общества играют
психологические исследования и службы,

      в том числе, проводимые на базе высших учебных заведений.

      Одним из средств проведения тестирований является данный сетевой ресурс.</h3>

  </div>

</body>
```

</html>

Приложение В Страница с анкетами

<head>

<title>Анкеты</title>

<meta charset="utf-8">

<link type="text/css" rel="stylesheet" href="styles/style.css">

</head>

<body>

<div id="wrap">

<div id="header">

<h1>Анкеты</h1>

<ul id="nav">

Главная

Анкеты

Авторизация

Регистрация

Обратная связь

</div>

<div id="intro">

<h3>Тут вы можете выбрать какую анкету или тест хотели бы пройти.</h3>

</div>

<div id="main">

<div id="content">

<p class="service-title first">Вопросник для выявления признаков вегетативных изменений</p>

<p>Описание 1 анкеты</p>

<p>НАЙМИГЕНСКИЙ тест</h1></p>

<p class="service-title first">Название 2 анкеты</p>

<p>Описание 2 анкеты</p>

<p>Индекс жизненного стиля</h1></p>

<p class="service-title first">Тест "Методика многостороннего исследования личности".</p>

<p>Описание 3 анкеты</p>

<p>Анкета</h1></p>

</div>

<div id="footer">

подвал

</div>

</div>

</body>

</html>