

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук  
Департамент программной инженерии

**СОГЛАСОВАНО**

Научный руководитель,  
старший преподаватель Департамента  
анализа данных и искусственного  
интеллекта

\_\_\_\_\_ А. А. Паринов  
«11» апреля 2022 г.

**УТВЕРЖДАЮ**

Академический руководитель  
ОП «Программная инженерия»  
профессор департамента программной  
инженерии, канд. техн. наук

\_\_\_\_\_ В. В. Шилов  
«\_\_» \_\_\_\_\_ 2022 г.

**СИМПТОМЧЕКЕР ДЛЯ ВЫЯВЛЕНИЯ КАРДИОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ  
С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ**

**Пояснительная записка**

**ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.05.11-01 81 01-1-ЛУ**

Подп. и дата	
Инв. № дубл.	
Взам. инв. №	
Подп. и дата	
Инв. № подл	RU.17701729.05.11-01 81 01-1-ЛУ У

Исполнитель  
студент группы БПИ192  
\_\_\_\_\_ / Д. С. Лямзин /  
«11» апреля 2022 г.

**Москва 2022**

**УТВЕРЖДЕН**

**RU.17701729.05.11-01 81 01-1-ЛУ**

**СИМПТОМЧЕКЕР ДЛЯ ВЫЯВЛЕНИЯ КАРДИОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ  
С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ**

**Пояснительная записка**

**RU.17701729.05.11-01 81 01-1**

**Листов 39**

<i>Подп. и дата</i>	
<i>Инв. № дубл.</i>	
<i>Взам. инв. №</i>	
<i>Подп. и дата</i>	
<i>Инв. № подл</i>	RU.17701729.05.11-01 81 01-1

**Москва 2022**

## СОДЕРЖАНИЕ

1.....	ВВЕДЕНИЕ	4
1.1. Наименование программы .....		4
1.2. Документы, на основании которых ведется разработка.....		4
2.....	НАЗНАЧЕНИЕ И ОБЛАСТЬ ПРИМЕНЕНИЯ	5
2.1. Функциональное назначение.....		5
2.2. Эксплуатационное назначение.....		5
3.....	ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ	7
3.1. Постановка задачи на разработку программы.....		7
3.2. Описание применяемых алгоритмов и функционирования программы .....		8
3.2.1.....	Общее описание процесса взаимодействия программы с пользователем	8
3.2.2.....	Общее описание процесса создания карточек с вопросами и алгоритма обработки пользовательских ответов чат-ботом	12
3.2.3.....	Описание алгоритма распознавания симптомов, указанных пользователем	13
3.2.4.....	Описание алгоритма генерации данных для обучения нейронной сети, отвечающей за выявление кардиологических заболеваний	16
3.2.5.....	Описание архитектуры нейронной сети, определяющей перечень кардиологических заболеваний, о наличии которых с наибольшей вероятностью свидетельствуют симптомы, указанные пользователем	18
3.2.6..	Описание процесса взаимодействия модуля, содержащего реализацию нейронной сети с остальной частью программы и алгоритма выбора уточняющего вопроса после получения результатов работы нейронной сети	20
3.3. Описание и обоснование выбора метода организации входных и выходных данных .....		21
3.4. Описание и обоснование выбора состава технических и программных средств.....		22
4.....	ОЖИДАЕМЫЕ ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ	23
4.1. Предполагаемая потребность.....		23
4.2. Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами .....		23

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

ИСТОЧНИКИ, ИСПОЛЬЗОВАННЫЕ ПРИ РАЗРАБОТКЕ.....	24
ПРИЛОЖЕНИЕ 1 .....	25
СПИСОК ТЕРМИНОВ .....	25
ПЕРЕЧЕНЬ СЕРДЕЧНО_СОСУДИСТЫХ ЗАБОЛЕВАНИЙ, ВЕРОЯТНОСТЬ НАЛИЧИЯ КОТОРЫХ МОЖЕТ БЫТЬ ОПРЕДЕЛЕНА СИМПТОМЧЕКЕРОМ.....	34
ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ .....	39

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

## 1. ВВЕДЕНИЕ

### 1.1. Наименование программы

Наименование программы – «Симптомчекер для выявления кардиологических заболеваний с использованием нейросетевых технологий».

Наименование программы на английском языке – «Symptom Checker for the Detection of Cardiac Diseases Based on Neural Networks ».

Краткое наименование программы - « CVSkeeper».

### 1.2. Документы, на основании которых ведется разработка

Программа выполняется в рамках курсового проекта «Симптомчекер для выявления кардиологических заболеваний с использованием нейросетевых технологий» [см. Приложение 1].

Основанием для разработки является учебный план подготовки бакалавров по направлению 09.03.04 “Программная инженерия” и утверждённая академическим руководителем тема курсового проекта.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

## 2. НАЗНАЧЕНИЕ И ОБЛАСТЬ ПРИМЕНЕНИЯ

### 2.1. Функциональное назначение

Основная функция программы - составление списка наиболее вероятных кардиологических заболеваний, о наличии которых свидетельствуют симптомы, указанные пользователем. Вместе с названием каждого заболевания в списке отображается вероятность его наличия и предоставляется ссылка на сайт, на котором опубликованы и медицинских исследований, позволяющих получить более полную информацию о том или ином диагнозе.

### 2.2. Эксплуатационное назначение

Программа будет использоваться для базовой оценки состояния здоровья человека, предварительного выявления кардиологических заболеваний и рисков их возникновения, а также оценки степени серьезности симптомов и принятия решения относительно срочности и целесообразности обращения к квалифицированным медицинским специалистам.

### 2.3. Краткая характеристика области применения

“Симптомчекер для выявления кардиологических заболеваний с использованием нейросетевых технологий” - это чат-бот, принимающий от пользователя информацию о беспокоящих его симптомах и возвращающий список кардиологических заболеваний, о которых они могут свидетельствовать. Разработка данной программы началась в 2021 году в рамках летней технологической практики на Факультете Компьютерных Наук НИУ ВШЭ. Тогда был создан прототип программы, способный распознавать менее 20 болезней и менее 100 симптомов. При подсчёте вероятности наличия у пользователя того или иного заболевания учитывалась лишь вычисленная вручную степень совпадения его характерных симптомов с симптомами, указанными пользователем, без определения наиболее и наименее важных и распространённых симптомов в рамках одной болезни.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

В рамках данной курсовой работы было значительно увеличено число распознаваемых симптомов и болезней, усовершенствован алгоритм подсчёта вероятности наличия того или иного заболевания (путём добавления к программе искусственно созданной нейронной сети) и проведена работа над улучшением качества распознавания симптомов.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

### 3. ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

#### 3.1. Постановка задачи на разработку программы

Разрабатываемая программа должна:

- 1) Получать и обрабатывать информацию о биологических характеристиках пользователя (пол, возраст, рост, вес и так далее);
- 2) Получать и обрабатывать информацию из “истории болезни” пользователя, потенциально связанную с кардиологическими заболеваниями и факторами риска их возникновения;
- 3) Получать и распознавать список симптомов, беспокоящих пользователя;
- 4) Определять список кардиологических заболеваний, о наличии которых с наибольшей вероятностью свидетельствуют симптомы, указанные пользователем;
- 5) Осуществлять сужение количества потенциальных болезней в списке до трёх или менее заболеваний путём уточнения и дополнения списка симптомов, указанных пользователем;
- 6) Возвращать в качестве выходных данных список потенциальных диагнозов пользователя со ссылками на базы данных, содержащие перечни статей, основанных на научных доказательствах и позволяющих собрать дополнительную информацию о том или ином заболевании по окончании уточнения симптомов [3];
- 7) Рассчитывать и выводить на экран вероятность наличия у пользователя того или иного кардиологического заболевания, выраженную вещественным числом от 0 до 1. Для более понятного и удобного представления вероятности пользователю перед выводом на экран допускается её предварительное умножение на 100 процентов и округление до двух знаков после запятой;
- 8) Возвращать в качестве выходных данных краткую информацию о болезнях, определённых симптомчекером как наиболее вероятные заболевания, о наличии которых свидетельствуют симптомы, указанные в качестве входных данных;
- 9) Выводить на экран рекомендации, касающиеся наблюдения за симптомами, проявление или исчезновение которых может помочь при

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и



дальнейшем определении диагноза пользователя во время консультации со специалистом.

### 3.2. Описание применяемых алгоритмов и функционирования программы

#### 3.2.1. Общее описание процесса взаимодействия программы с пользователем

Для начала работы с симптомчекером пользователь должен отправить ему любое сообщение, после чего чат-бот присылает карточку с просьбой об указании возраста и дожидается ответа, который необходимо ввести в специальное поле и отправить при помощи нажатия на кнопку “Далее” (рис. 1). Затем в диалоге с программой пользователю необходимо аналогичным образом указать свой пол, рост и вес, после чего начнут появляться карточки с вопросами, позволяющими выявить наличие у пользователя рисков возникновения сердечно-сосудистых заболеваний и получить некоторые данные о его “истории болезни”, способные повлиять на диагностику (наличие уже обнаруженных кардиологических заболеваний, хронических заболеваний, способных давать осложнения на сердце, родственников с сердечно-сосудистыми заболеваниями, последствий перенесённых инфекционных заболеваний или сезонной аллергии, а также качество питания, уровень физической активности, наличие тяги к курению и употреблению алкоголя). Факторы, повышающие вероятность обнаружения тех или иных болезней, не используются в дальнейшей работе симптомчекера, но - так же, как и все входные данные, поступающие со стороны пользователя - сохраняются в файл “answers.txt” в формате “<id пользователя> <id вопроса> <текст ответа>” и могут быть применены при обучении нейронной сети и усовершенствовании качества распознавания болезней в обозримом будущем.

Сколько вам лет?

40

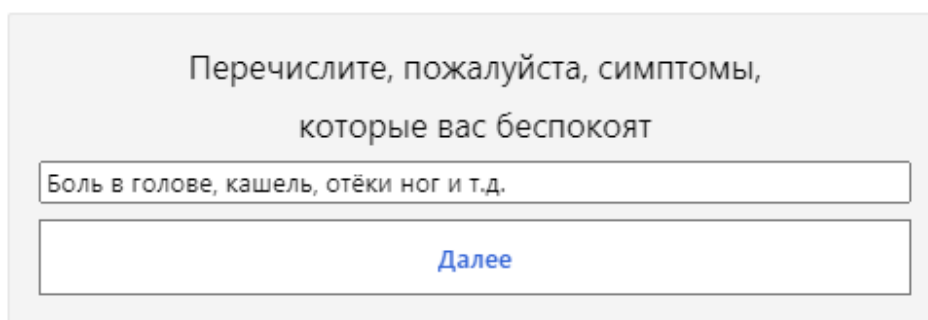
Далее

Just now

Рисунок 1. Первый вопрос, задаваемый чат-ботом.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

Следующий шаг взаимодействия пользователя с программой - ввод беспокоящих его симптомов (рис. 2). После отправки симптомов и их считывания чат-ботом (см. п. 3.2.3.) пользователь может убедиться, в том, что программа правильно распознала его намерения и при необходимости добавить или удалить недостающие или лишние данные (рис. 3). В конечном итоге информация о симптомах, беспокоящих пользователя, передаётся в качестве входных данных нейронной сети, осуществляющей подсчёт вероятности наличия у пользователя того или иного заболевания сердечно-сосудистой системы и возвращающей на выходе шесть вещественных чисел: вычисленные вероятности для трёх болезней, для которых (по её мнению) представленные симптомы являются наиболее характерными, а также идентификаторы соответствующих им кардиологических заболеваний.



Перечислите, пожалуйста, симптомы,  
которые вас беспокоят

Боль в голове, кашель, отёки ног и т.д.

Далее

Рисунок 2. Пример просьбы о вводе симптомов.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

Мне удалось распознать следующие симптомы:

головная боль,  
кашель,  
отёки нижних конечностей.

Проверьте, всё ли правильно, и попробуйте  
добавить или убрать симптомы,  
если что-то не так.

Добавить симптомы

Убрать симптомы

Далее

Рисунок 3. Пример просьбы о проверке и редактировании симптомов.

После обработки результатов, выданных нейросетью, “CVSkeeper” начинает задавать пользователю уточняющие вопросы (рис. 4). Каждый уточняющий вопрос имеет три возможных варианта ответа (“да”, “нет”, “не знаю”) и сообщает чат-боту информацию о присутствии или отсутствии у пользователя ровно одного симптома (таким образом, число карточек с уточняющими вопросами совпадает с количеством симптомов, которое способна распознать программа). После получения ответа на каждый уточняющий вопрос система предлагает пользователю подтвердить или опровергнуть свои показания (рис. 5) и после того, как тот соглашается на продолжение работы или отменяет предыдущее действие, давая новый ответ на последний вопрос, вновь обращается к нейросети для получения новых вероятностей наличия заболеваний с уже скорректированными симптомами. Уточняющие вопросы задаются пользователю до тех пор, пока не произойдёт одно из трёх событий: 1) нейросеть три раза подряд вернёт одинаковые списки из трёх наиболее вероятных болезней пользователя после уточнения каких-либо симптомов, 2) вероятность наличия одного и того же заболевания, вычисляемая нейронной сетью после уточнения какого-либо симптома, три раза подряд превысит 0.9, 3) количество уточняющих вопросов, заданных пользователю, достигнет двадцати.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

Отмечаете ли вы резкий беспричинный  
рост массы тела в последние месяцы?

Да

Нет

Не знаю

Today

Рисунок 4. Пример уточняющего вопроса, задаваемого симптомчекером.

Вы уверены?

Today

Да

Today

Рисунок 5. Пример просьбы о подтверждении ответа на уточняющий вопрос и реакции пользователя на неё.

После отправки пользователю сообщения об окончании этапа уточнения симптомов чат-бот выводит информацию о результатах последней итерации работы нейронной сети (рис. 6), а также отправляет краткую информацию по каждому представленному заболеванию, содержащую основную характеристику патологии, ссылку на перечень статей по диагнозу, опубликованных в медицинской базе данных PubMed и перечисление наиболее характерных симптомов болезней, на которые следует обращать особое внимание при принятии решения об обращении к терапевту или к профильному специалисту.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

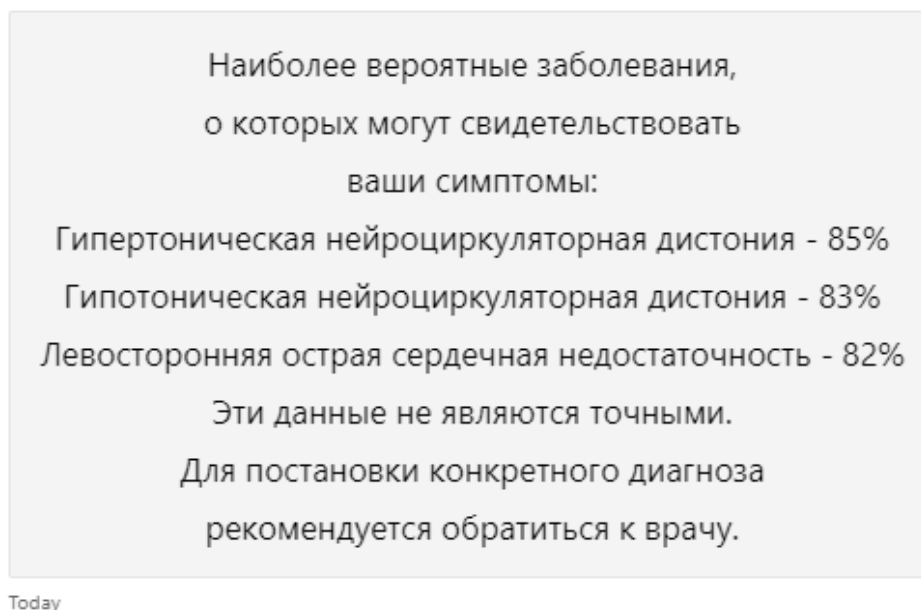


Рисунок 6. Пример сообщения о результатах работы программы.

### 3.2.2. Общее описание процесса создания карточек с вопросами и алгоритма обработки пользовательских ответов чат-ботом

Разработка симптомчекера велась при помощи сервиса “Bot Framework” [2]. Процесс взаимодействия пользователя с чат-ботом с точки зрения программы был организован таким образом, чтобы система в асинхронном режиме дожидалась активности пользователя, считывала содержимое отправленных сообщений и реагировала на них в зависимости от значения счётчика, сохраняемого в переменной “message\_number”, и изменяющегося после ответов пользователей на любые вопросы.

Формирование содержимого карточек с вопросами осуществляется благодаря использованию типа данных AdaptiveCards, поддерживаемого “Bot Framework” и позволяющего создавать сообщения с большим количеством вложений и размещать в них текстовые данные, поля для ввода, кнопки, изображения и другие распространённые элементы графического интерфейса на усмотрение разработчика (рис. 7). Асинхронная отправка выходных данных осуществляется через ещё один класс “Bot Framework” под названием “MessageFactory” с передачей ссылки на содержимое сообщения и токена, запрашиваемого классом “Task” для получения доступа к отмене тех или иных задач.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

```

if (number == 50)
{
    AdaptiveCard card = new AdaptiveCard();
    card.Body.Add(new AdaptiveTextBlock
    {
        Text = "Ваш кашель можно назвать сухим?",
        Size = AdaptiveTextSize.Medium,
        HorizontalAlignment = AdaptiveHorizontalAlignment.Center
    });
    card.Actions.Add(new AdaptiveSubmitAction
    {
        Title = "Да",
        Data = new AttachmentData().Type = "Да"
    });
    card.Actions.Add(new AdaptiveSubmitAction
    {
        Title = "Нет",
        Data = new AttachmentData().Type = "Нет"
    });
    card.Actions.Add(new AdaptiveSubmitAction
    {
        Title = "Не знаю",
        Data = new AttachmentData().Type = "Не знаю"
    });
    Attachment attachment = new Attachment()
    {
        ContentType = AdaptiveCard.ContentType,
        Content = card
    };
    await turnContext.SendActivityAsync(MessageFactory.Attachment(attachment), cancellationToken);
}

```

Рисунок 7. Пример устройства карточки с вопросом с точки зрения кода.

### 3.2.3. Описание алгоритма распознавания симптомов, указанных пользователем

Перед началом курсовой работы модуль программы, отвечающий за определение симптомов и жалоб пользователя, указываемых в строке, отправляемой симптомчекеру в качестве входных данных на стадии выявления симптомов, распознавал симптомы исключительно по точному совпадению описывающих их фраз или частей фраз, со словами или частями слов, введёнными пользователем (рис. 8). Входная строка разбивалась на несколько “интентов” (в качестве разделителя выступала запятая), и каждый “интент” обрабатывался таким образом, чтобы в нём не осталось лишних пробелов, знаков препинания и иных символов, затрудняющих распознавание симптомов, после чего производились попытки выделения из каждой фразы отдельных смысловых узлов, обозначающих вид недуга (например, “боль”, “онемение”, “жжение”, покраснение” и т. д.), локализацию его проявления (“нижние конечности”, “верхние конечности”, “лицо”, “кожа”, “шея” и т. д.), а также характер наблюдаемых изменений (“увеличение”, “учащение”, “замедление”, “потеря” и т.д.), и комбинации смысловых узлов объединялись в симптомы (рис. 9).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

```

Ссылка: 14
public int RecognizeBurning(string intent)
{
    List<string> burningExpressions = new List<string> { "жжет", "жжен", "жгуч" };
    foreach (string expression in burningExpressions)
    {
        if (intent.Contains(expression))
        {
            return 1;
        }
    }
    return 0;
}

```

Рисунок 8. Пример распознавания симптома по частям слов.

```

if (RecognizeGums(intent) == 1 && RecognizeBlood(intent) == 1)
{
    symptomsList.Add("кровоточивость дёсен");
}

if (RecognizeSputum(intent) == 1 && RecognizeBlood(intent) == 1)
{
    symptomsList.Add("кровь в мокроте");
}

if (RecognizeHearing(intent) == 1 && (RecognizeImpairment(intent) == 1 || RecognizeDecrease(intent) == 1))
{
    symptomsList.Add("нарушение слуха");
}

if (RecognizeFainting(intent) == 1 || RecognizeConscience(intent) == 1)
{
    symptomsList.Add("обморочное состояние");
}

if (RecognizeBreath(intent) == 1)
{
    symptomsList.Add("одышка");
}

```

Рисунок 9. Примеры объединения комбинаций смысловых узлов в симптомы.

В рамках данной разработки было увеличено число симптомов и смысловых узлов, которые способен распознавать чат-бот [см. Приложение 2], произведено расширение списка фраз и отрывков слов, которые можно расценивать как отдельные смысловые узлы, и усовершенствован процесс препроцессинга данных путём обработки возможных орфографических ошибок, допущенных пользователем при вводе симптомов при помощи сервиса “Яндекс.Спеллер” [8] (рис. 10). Для предотвращения ситуаций, в которых программа не может распознать ту или иную жалобу человека из-за того, что в качестве входной фразы было использовано правильно написанное слово, незначительно отличающееся от какого-либо слова, воспринимаемого чат-ботом как указание на определённый симптом (например, с изменённым окончанием), была добавлена функция распознавания фраз, для которых кратчайшее расстояние Левенштейна до какого-либо симптома не превышает трети их длины (либо количества символов во фразе,

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

описывающей симптом; в обоих случаях длина рассматриваемых строк считается после удаления из неё всех пробелов и знаков препинания). Для подсчёта редакционного расстояния используется алгоритм Вагнера-Фишера [5] (рис. 11), принимающий на вход строку обработанного “интента”, отдельные слова которого были предварительно переставлены таким образом, чтобы они шли в лексикографическом порядке (чтобы, например, фраза “носовые кровотечения” не была расположена слишком далеко от фразы “кровотечения из носа”), а также аналогично обработанную и “отсортированную” строку описания симптома.

```

Ссылка: 2
public void RecognizeIntents()
{
    var speller = new SpellerService();

    foreach (string intent in intents)
    {
        var new_intent = speller.CheckTextAsync(intent).Result[0].Steers[0];

        if (RecognizeSkinProblems(new_intent) != "")
        {
            symptomsList.Add(RecognizeThroatProblems(new_intent));
        }

        if (RecognizeThroatProblems(new_intent) != "")
        {
            symptomsList.Add(RecognizeThroatProblems(new_intent));
        }

        if (RecognizeChestProblems(new_intent) != "")
        {
            symptomsList.Add(RecognizeChestProblems(new_intent));
        }
    }
}

```

Рисунок 10. Код, иллюстрирующий один из способов работы с сервисом “Яндекс.Спеллер” для исправления орфографических ошибок.

```

int levensteinInstruction(String s1, String s2, int InsertCost, int DeleteCost, int ReplaceCost):
    D[0][0] = 0
    for j = 1 to N
        D[0][j] = D[0][j - 1] + InsertCost
    for i = 1 to M
        D[i][0] = D[i - 1][0] + DeleteCost
        for j = 1 to N
            if S1[i] != S2[j]
                D[i][j] = min(D[i - 1][j] + DeleteCost,
                             D[i][j - 1] + InsertCost,
                             D[i - 1][j - 1] + ReplaceCost)
            else
                D[i][j] = D[i - 1][j - 1]
    return D[M][N]

```

Рисунок 11. Псевдокод, описывающий реализацию алгоритма Вагнера-Фишера.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и



### 3.2.4. Описание алгоритма генерации данных для обучения нейронной сети, отвечающей за выявление кардиологических заболеваний

Поскольку ни один из найденных во время анализа предметной области датасетов, содержащих медицинские данные, не включал в себя всех симптомов и диагнозов, с которыми уже работал симптомчекер до начала курсовой работы, в качестве материала для обучения нейронной сети был использован датасет, состоящий из 40000 записей, сгенерированных программным алгоритмом на основании информации, полученной при консультации с профессиональными специалистами, имеющими медицинское образование.

Каждая строка датасета (рис. 12) состоит из 227 вещественных чисел из диапазона от 0 до 1 включительно, разделённых пробелами. Первые 226 чисел обозначают каждый из симптомов, описанных в Приложении 2, а последнее число обозначает диагноз, который должен быть поставлен пациенту с соответствующими симптомами по мнению врача. В качестве возможных диагнозов выступают 40 различных сердечно-сосудистых заболеваний [см. Приложение 3], и каждому заболеванию соответствуют 1000 строк датасета. Изначально все строки датасета были заполнены нулями. Затем был произведён анализ частоты встречаемости каждого симптома у пациентов, страдающих каждым заболеванием (одна часть данных была взята из исследований, опубликованных в базе данных “PubMed”, другая основана на личном опыте специалиста, проводившего консультацию), и, согласно найденным частотам, происходило распределение положительных вещественных чисел по датасету таким образом, чтобы для набора строк, соответствующих одному заболеванию, доля ненулевых значений на тех позициях строк, которые соответствуют одному симптому, была равна вероятности проявления рассматриваемого симптома при рассматриваемом заболевании (таким образом, если было установлено, что в среднем у 80% пациентов, страдающих стенокардией, наблюдается боль за грудиной или в области сердца, то в 800 из 1000 строк датасета, описывающих данные для стенокардии, числа, описывающие симптом “боль за грудиной или в области сердца”, будут ненулевыми). Значение каждого ненулевого элемента датасета зависит от числа  $N$  - примерной вероятности того, что утверждение пользователя о наличии или отсутствии у него того или иного симптома окажется истинным - и является случайным числом, находящимся на отрезке от 0.5 до  $N + (N - 0.5)$ , если  $N \leq 0.75$  и на отрезке от  $N - (1 - N)$  до 1, если  $N > 0.75$ . Данная мера позволяет обучить нейросеть таким образом, чтобы результаты её работы учитывали степень неуверенности пользователя программы в своих ответах и чтобы нейросеть выдавала более точные

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

результаты для ситуации, когда пользователь не может самостоятельно обнаружить симптом в домашних условиях без специализированного обследования.

0.9330449782196171 0.8347412659524456 0.6782204926676438 0 0 0 0 0.9102477381013698 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.7664347136084972 0.7470695950006316  
0.9896846672794753 0.8731028463172723 0 0 0 0 0.999857231858318 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.5716912008259077 0.600859335354782 0.7378897296291134  
0.9287348170165334 0.5748554122934102 0.5295881952146231 0 0 0 0 0.9620521291043925 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6928565257797294 0.8251716607142789  
0.917186877259073 0.6240632399502732 0.6520916683950755 0 0 0 0 0.9573934521830016 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6131724913898187 0.7416912228401239  
0.936827610269424 0.7452769887634323 0.6620490041805315 0 0 0 0 0.99072582285359 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.574777314104135 0.7790871972315538  
0.935852269079129 0.9566153772787239 0 0 0 0 0.9077584066831258 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.506250839485126 0.5806923334809516 0.7689973812353934  
0.9256607683557148 0.7531132013986042 0.9807472910364768 0 0 0 0 0.9056819483233099 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.7874403594218725 0.997459345087820  
0.9261871113186801 0.5564359828837937 0.6594853112002959 0 0 0 0 0.9362691062238755 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.5552808129754244 0.7840752900274928  
0.9285605896702073 0.7676274518634882 0.5159231019787429 0 0 0 0 0.9428550536994125 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6597823421903442 0.9239149785541622  
0.9138351149306504 0.6051683093660547 0.5810313662183538 0 0 0 0 0.9031184571106944 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6890834680736664 0.7911246877624197  
0.9723389390827768 0.5380733201442631 0.893243560621336 0 0 0 0 0.98125435222447 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.5532018922614417 0.710161263958837  
0.9388724005819671 0.9408076068795617 0 0 0 0 0.9571653227749377 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6808502585360969 0.6836302952513815 0.722622102700226  
0.9534151880831925 0.6757748412319597 0.5742750391330642 0 0 0 0 0.9557083251358862 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6470030906510096 0.7850695883734005  
0.9487227896543107 0.5094874665712404 0.9461983623898831 0 0 0 0 0.9213103855738637 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6367307017777729 0.7066311344580146  
0.9036033727425513 0.8137470250670256 0.965687665749917 0 0 0 0 0.9711730138076451 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.8205891886527845 0.97988453023130  
0.9152686175288486 0.7132827382785327 0.7390073199341751 0 0 0 0 0.9140320815635526 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6803326166985354 0.7065983558670859  
0.6151207563915674 0.6335915630117417 0 0 0 0 0.9578676137003949 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6072859798483842 0.7463953780011853 0.9509151856653241  
0.972731593520604 0.5919931911129637 0.958549477107445 0 0 0 0 0.966588726621119 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.5926004675658716 0.941290481281054  
0.9490201228840788 0.8495710904288658 0.9680254803763144 0 0 0 0 0.9336238577346676 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.7968051381952175 0.86460902589770  
0.9252002373329047 0.7008378679275208 0.7137647883010934 0 0 0 0 0.92497266904477 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.590741644239574 0.7119960021124979  
0.9411390220650975 0.5648767376393782 0.6211352780143657 0 0 0 0 0.9170392680555914 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.5877868968898962 0.8534749006387851  
0.9434449882894725 0.81923241148298 0.5004186284739790 0 0 0 0 0.9187245316990904 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6042495433694253 0.8481899357809529  
0.9418770213183342 0.7095469601705089 0.5180510947933368 0 0 0 0 0.9590840941170063 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.636642097722651 0.9124972466293138  
0.926669837920558 0.8273489227625517 0.9376710124808458 0 0 0 0 0.9992844228731632 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.86848906221516382 0.9417987953240  
0.9808793799714385 0.9160901130535952 0 0 0 0 0.9041315365784468 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.66139412221487725 0.6160648280012746 0.7562306180293395  
0.9023424749601955 0.6124952958028122 0.6613632453161982 0 0 0 0 0.9039695551444129 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.66258576610266 0.882484475933493  
0.9071729192701669 0.9558170536305618 0 0 0 0 0.9448123331450943 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.501618642357029 0.8206863943547527 0.7738919840110075  
0.9237414716358765 0.6330675425219593 0.5540329101039007 0 0 0 0 0.947610368797047 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.6907109926700251 0.9256509095780075

Рисунок 12. Фрагмент датасета, на котором представлено отображение симптомов стенокардии

Для того, чтобы содержимое датасета было максимально приближено к реальным данным, заполнение строк информацией происходило посимптомно от наиболее распространённого симптома к наиболее редкому. Для улучшения результатов работы нейросети в ситуации, когда симптомы, введённые пользователем, в точности совпадают с симптомами одного из заболеваний, X из 1000 строк каждого раздела датасета, относящегося к одной болезни, содержали в себе ненулевые элементы на каждой позиции, соответствующий характерному для неё симптому (X - округлённая до целого числа вероятность проявления всех характерных симптомов заболевания одновременно, равная произведению отдельных вероятностей; если X оказывается равным нулю, ему автоматически присваивается значение 1). Остальные строки заполнялись значениями так, чтобы выполнялись все вышеописанные условия и при этом в каждый момент времени новое ненулевое значение появлялось в той строке, сумма элементов которой была наименьшей среди всех строк, соответствующих обрабатываемому заболеванию [см. Приложение 4].

После формирования датасета было произведено его изучение, и в рамках очередной консультации с медицинским работником была произведена верификация и валидация полученных данных, по итогам которой были решены все конфликтные ситуации, связанные с присутствием в одной строке нескольких ненулевых значений, соответствующих взаимоисключающим симптомам, и было принято решение о том, что реалистичность датасета позволяет разработчику симптомчекера использовать его для

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

обучения нейронной сети. Поскольку в реальности при дифференциальной диагностике для постановки диагноза в спорных случаях учитываются не только симптомы, но и распространённость заболеваний, которым они соответствуют, в качестве ожидаемого значения, свидетельствующего о полном соответствии указанных симптомов тому или иному заболеванию (последний столбец датасета), будут использоваться вещественные числа от 0 до 1, причём их значение заранее подбирается для каждого заболевания таким образом, чтобы на единичной шкале порядок точек, соответствующих сорока болезням, совпадал с порядком следования болезней в списке, указанном в Приложении 3 (они расположены таким образом, чтобы на соседних позициях стояли диагнозы, наиболее близкие по наблюдаемым симптомам друг к другу), а соотношение длин числовых промежутков, центрами которых являются данные точки, совпадали с выявленным соотношением распространённости заболеваний.

### **3.2.5. Описание архитектуры нейронной сети, определяющей перечень кардиологических заболеваний, о наличии которых с наибольшей вероятностью свидетельствуют симптомы, указанные пользователем**

Построенная модель нейронной сети является моделью глубокого распространения (многослойный перцептрон) и состоит из пяти слоёв: один входной слой, один выходной слой и три скрытых слоя. Входной слой состоит из 226 элементов и принимает на вход вектор элементов датасета, первый и второй скрытые слои состоят из 64 нейронов, третий скрытый слой состоит из 16 нейронов и, наконец, выходной слой состоит из одного выходного нейрона, значение которого будет является предсказанием, выданным моделью и будет сравниваться с соответствующими значениями из тестовых данных (рис. 13). В качестве функции активации для всех нейронов была выбрана функция ReLu, однако для выходного нейрона вместо неё используется функция сигмоиды, поскольку на выходе необходимо получить число от 0 до 1 [10].

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

```

Model: "sequential"
_____
Layer (type)                Output Shape                Param #
-----
dense (Dense)                (None, 64)                  14528
dense_1 (Dense)              (None, 64)                  4160
dense_2 (Dense)              (None, 64)                  4160
dense_3 (Dense)              (None, 16)                  1040
dense_4 (Dense)              (None, 1)                   17
_____
Total params: 23,905
Trainable params: 23,905
Non-trainable params: 0

```

Рисунок 13. Схематичное изображение архитектуры нейронной сети.

Вероятности наличия у пользователя определённого заболевания не совпадают со значением выходного нейрона и рассчитываются на его основе по следующей формуле:

$$A = 0.5 + 0.5 * \max(0, 1 - B / C),$$

где A - вероятность наличия заболевания,

B - модуль разности между предсказанием нейросети и ожидаемым значением для рассматриваемой болезни из датасета,

C - абсолютная длина отрезка, “отведённого” рассматриваемому заболеванию методом, описанным в п. 3.2.4.

Благодаря данной формуле каждая из трёх найденных вероятностей будет не меньше 0.5, а распределение вероятностей между двумя заболеваниями, которые нейросеть посчитала наиболее соответствующими входным симптомам будет пропорционально разности модулей ожидаемого и фактического значения для каждой из них (таким образом, при одинаковой близости значения выходного нейрона к ожидаемым значениям для двух различных заболеваний, определённые программой вероятности наличия двух этих заболеваний будут одинаковыми).

Обучение модели нейросети производилось на 1000 эпохах, в качестве функции потерь использовалась средняя абсолютная ошибка (MAE) [6]. В качестве оптимизатора был выбран оптимизатор “Adam”.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

Весь код, отвечающий за генерацию датасета, а также за построение и обучение модели нейронной сети, был написан на языке программирования Python, в то время как часть симптомчекера, отвечающая за непосредственное взаимодействие с пользователем, разрабатывалась на C#. Для передачи данных от одной части программы к другой была использована страница симптомчекера “CVSkeeper” в социальной сети “ВКонтакте” [1].

При помощи технологий, предоставляемых API “ВКонтакте” для двух разных языков программирования [4], в программный модуль на C# была добавлена функция отправки строки, содержащей 226 вещественных чисел, соответствующих каждому из потенциально распознаваемых симптомов (симптомы, изначально указанные или подтверждённые пользователем за счёт предоставления утвердительного ответа на уточняющий вопрос, кодировались числом 1, отсутствующие или неподтверждённые симптомы - числом 0, а симптомы, при уточнении которых пользователь дал ответ “не знаю” - числом 0.5) (рис. 14).

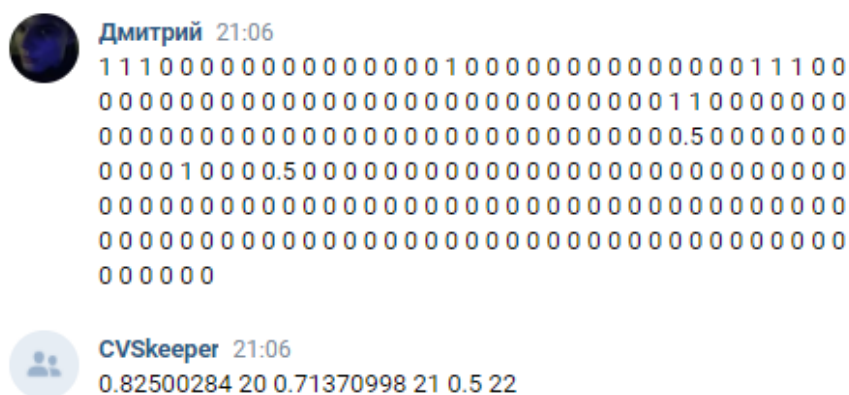


Рисунок 14. Пример взаимодействия основной части программы с частью, связанной с работой нейронной сети.

Передача результатов работы нейронной сети производилась аналогичным образом. Отправляемое сообщение имело формат “<вероятность наличия болезни 1> <идентификатор болезни 1> <вероятность наличия болезни 2> <идентификатор болезни 2> <вероятность наличия болезни 3> <идентификатор болезни 3>”. Идентификатор болезни -

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

целое число, соответствующее порядковому номеру заболевания в списке, указанном в Приложении 3.

После получения выходных данных от нейросети чат-бот составляет перечень ещё не уточнявшихся симптомов, наличие которых характерно только для двух или только для одного из заболеваний, считающихся наиболее вероятными потенциальными диагнозами пользователя в текущий момент времени, и выбирает из него симптом таким образом, чтобы произведение средней относительной распространённости заболеваний, для которых симптом характерен (среди трёх анализируемых заболеваний), и вероятности достоверности информации о симптоме, получаемой от пользователя (указанной в Приложении 2) оказалось максимально возможным.

### **3.3. Описание и обоснование выбора метода организации входных и выходных данных**

Выходные данные должны передаваться пользователю в виде сообщений, содержащих информацию о результатах работы программы, просьбы о вводе данных или вопросы, ответы на которые необходимо передавать симптомчекеру при помощи текстовых сообщений или за счёт нажатия на кнопки с текстом, обозначающего выбор пользователем одного из нескольких действий, являющихся частью его взаимодействия с программой

Выбор данного метода организации входных и выходных данных обусловлен тем, что интерфейс чат-бота может встраиваться интерфейсы мессенджеров, социальных сетей, веб-сервисов или приложений. Подавляющее большинство из них поддерживает отправку текстовых сообщений и ввод данных путём нажатия на кнопки, однако некоторые другие возможности, предоставляемые сервисом “Bot Framework”, в них могут оказаться недоступными, так что добавление дополнительных методов организации входных и выходных данных было бы избыточным, а удаление или изменение уже существующих методов - нецелесообразным, поскольку без ввода текстовых сообщений программа не могла бы получить от пользователя всю необходимую ей информацию, а отсутствие возможности отправки ответа и выбора подходящего варианта нажатием на кнопку сделало бы использование программы менее удобным и более утомительным.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

### 3.4. Описание и обоснование выбора состава технических и программных средств

Основная цель создания симптомчекера предполагает его использование на различных интернет-платформах, и данная задача легко решается при помощи разработки части программного продукта на “Bot Framework”, позволяющего осуществлять быструю интеграцию чат-ботов не только в мессенджеры, но и во множество других сайтов и сервисов.

Нейронная сеть создавалась и обучалась на языке Python, так как именно на этом языке написаны наиболее популярные и многофункциональные библиотеки для машинного обучения и анализа данных (“Keras”, “Tensorflow”, “PyTorch”, “Pandas” и т.д.).

Использование данных технологий приводит к тому, что пользователю для работы с программой достаточно иметь лишь электронное устройство с клавиатурой (без неё невозможна отправка входных данных таким образом, чтобы их смог распознать симптомчекер), экраном (без него невозможно просматривать выходные данные, от реакции пользователя на которые зависят сценарии работы чат-бота) и свободным доступом в Интернет (без выполнения данного условия пользователь скорее всего не сможет подключиться к чат-боту).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

#### 4. ОЖИДАЕМЫЕ ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

##### 4.1. Предполагаемая потребность

Данный программный продукт будет интересен людям, заботящимся о состоянии своего здоровья или о состоянии здоровья своих близких, желающим проверить серьёзность беспокоящих их симптомов для принятия решения о необходимости обращения за медицинской помощью. Его целевой аудиторией могут быть люди любого возраста и пола. Интерфейс симптомчекера может как встраиваться в интерфейс обычного мессенджера (например, Telegram), так и интегрироваться в интерфейс различных веб-сайтов. Особенно актуальным и полезным было бы использование чат-бота на интернет-ресурсах, посвященных профилактике сердечно-сосудистых заболеваний.

##### 4.2. Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами

Сравнительный анализ создаваемого продукта и разработанных ранее программ, обладающих схожим функционалом и решающих схожие задачи, позволил выявить целый ряд сервисов-аналогов, созданных в России и за рубежом и использующих нейросетевые технологии для выявления различных заболеваний по указываемым симптомам (в частности, “Тест на симптомы” от компании “Сбер” [9] и симптомчекер “Helzy” [7]). В отличие от многих других программ, позволяющих проверить наличие или отсутствие симптомов только одного заболевания или работающих со всеми наиболее распространёнными болезнями, данная разработка будет сфокусирована на заболеваниях одного типа и будет направлена на то, чтобы пользователь мог получить как можно больше информации о его потенциальных диагнозах и о различиях между ними, а также имел возможность ознакомиться со списком симптомов, за развитием которых ему стоит наблюдать в течение некоторого времени после взаимодействия с симптомчекером.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и



## ИСТОЧНИКИ, ИСПОЛЬЗОВАННЫЕ ПРИ РАЗРАБОТКЕ

- 1) CVSSkeeper [Электронный ресурс]// URL: <https://vk.com/public212606321>. Режим доступа: свободный. Дата обращения: 11.04.2022.
- 2) Microsoft Bot Framework [Электронный ресурс]// URL: <https://dev.botframework.com/>. Режим доступа: свободный. Дата обращения: 15.06.2021.
- 3) PubMed [Электронный ресурс]// URL: <https://pubmed.ncbi.nlm.nih.gov/>. Режим доступа: свободный. Дата обращения: 06.12.2021.
- 4) VKNET Вконтакте API для .NET (C#) [Электронный ресурс]// URL: <https://vknet.github.io/vk/>. Режим доступа: свободный. Дата обращения: 06.12.2021.
- 5) Задача о редакционном расстоянии, алгоритм Вагнера-Фишера [Электронный ресурс]// URL: [https://neerc.ifmo.ru/wiki/index.php?title=Задача\\_о\\_редакционном\\_расстоянии,\\_алгоритм\\_В\\_агнера-Фишера](https://neerc.ifmo.ru/wiki/index.php?title=Задача_о_редакционном_расстоянии,_алгоритм_В_агнера-Фишера). Режим доступа: свободный. Дата обращения: 13.03.2022.
- 6) Индикаторы оценки регрессии MSE, RMSE, MAE, R-Squared [Электронный ресурс]// URL: <https://russianblogs.com/article/7295554016/>. Режим доступа: свободный. Дата обращения: 03.10.2021
- 7) Определить диагноз по симптомам онлайн | Симптомчекер Helzy [Электронный ресурс]// URL: <https://helzy.ru/>. Режим доступа: свободный. Дата обращения: 06.07.2021.
- 8) Спеллер - Технологии Яндекса [Электронный ресурс]// URL: <https://yandex.ru/dev/speller/>. Режим доступа: свободный. Дата обращения: 13.03.2022.
- 9) Тест на симптомы [Электронный ресурс]// URL: <https://www.sberbank.ru/promo/eco/covid/>. Режим доступа: свободный. Дата обращения: 05.07.2021.
- 10) Функции активации нейросети: сигмоида, линейная, ступенчатая, ReLu, tahn [Электронный ресурс]// URL: <https://neurohive.io/ru/osnovy-data-science/activation-functions/>. Режим доступа: свободный. Дата обращения: 20.11.2021.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

## СПИСОК ТЕРМИНОВ

Ниже приведён список терминов, необходимых для ознакомления:

- 1) **API (application programming interface)** - описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой.
- 2) **PubMed** - англоязычная текстовая база данных медицинских и биологических публикаций.
- 3) **Датасет** - набор очищенных данных, пригодных для обработки алгоритмами машинного обучения.
- 4) **Нейросетевые технологии** - комплекс информационных технологий, основанных на применении искусственных нейронных сетей.
- 5) **Нейросеть (Искусственная нейронная сеть, ИНС)** - математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей - сетей нервных клеток живого организма.
- 6) **Обучение нейронной сети** - процесс подбора входных весов для каждого нейрона таким образом, чтобы на выходе получить результат максимально соответствующий ожиданиям.
- 7) **Симптомчекер** — сервис, позволяющий поставить диагноз на основании симптомов, выявленных у пациента.
- 8) **Чат-бот** - виртуальный собеседник, программа, которая создана для имитации поведения человека при общении с одним или несколькими собеседниками.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

## ПРИЛОЖЕНИЕ 2

**СПИСОК СИМПТОМОВ КАРДИОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ,  
РАСПОЗНАВАЕМЫХ ПРОГРАММОЙ (С ПРИБЛИЗИТЕЛЬНЫМИ  
ПРОЦЕНТАМИ ВЕРОЯТНОСТИ ТОГО, ЧТО ИНФОРМАЦИЯ, СООБЩАЕМАЯ  
ПОЛЬЗОВАТЕЛЕМ О СИМПТОМЕ, ОКАЖЕТСЯ ДОСТОВЕРНОЙ)**

1. Боль за грудиной или в области сердца (95)
2. Боль за грудиной или в области сердца возникает внезапно (75)
3. Боль за грудиной или в области сердца возникает при физической активности (75)
4. Колющая боль за грудиной или в области сердца (65)
5. Тянущая боль за грудиной или в области сердца (65)
6. Давящая боль за грудиной или в области сердца (65)
7. Тупая боль за грудиной или в области сердца (65)
8. Боль в животе (95)
9. Пульсация в районе живота (80)
10. Боль в пояснице (85)
11. Боль в шее (90)
12. Боль при жевании (95)
13. Боль в глазах (90)
14. Боль в мышцах (85)
15. Боль в мышцах ног (85)
16. Боль в икроножных мышцах (85)
17. Боль в мышцах рук (85)
18. Боль в суставах (85)
19. Боль в пальцах (90)
20. Боль в нижних конечностях (95)
21. Ноющая боль в нижних конечностях (60)
22. Боль в нижних конечностях при ходьбе (80)
23. Боль в пальцах ног (85)
24. Боль в верхних конечностях (95)
25. Боль в пальцах рук (85)
26. Учащённое сердцебиение (70)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

- 27. Замедленное сердцебиение (55)
- 28. Ощущение перебоев в работе сердца (60)
- 29. Слабость пульса (55)
- 30. Ощущение стеснения (сдавливания) в груди (65)
- 31. Ощущение жжения в груди (70)
- 32. Иррадиация боли из грудной клетки в другие части тела (85)
- 33. Одышка (85)
- 34. Одышка появляется внезапно (65)
- 35. Одышка появляется при физической активности (75)
- 36. Ортопноэ (95)
- 37. Одышка усиливается в положении сидя или стоя (95)
- 38. Учащение дыхания (80)
- 39. Замедление дыхания (60)
- 40. Тошнота (90)
- 41. Рвота (100)
- 42. Чувство тревоги (50)
- 43. Слабость (65)
- 44. Чувство усталости и быстрая утомляемость (50)
- 45. Онемение лица (90)
- 46. Онемение пальцев (85)
- 47. Онемение верхних конечностей (85)
- 48. Онемение верхних конечностей с одной стороны (95)
- 49. Онемение пальцев рук (85)
- 50. Онемение нижних конечностей (85)
- 51. Онемение нижних конечностей с одной стороны (95)
- 52. Онемение пальцев ног (85)
- 53. Отёк лёгких (60)
- 54. Отёк лица (85)
- 55. Отёки на коже головы (75)
- 56. Отёк шеи (80)
- 57. Отёки нижних конечностей (90)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

- 58. Отеки верхних конечностей (90)
- 59. Повышенное потоотделение (80)
- 60. "Холодный пот" (70)
- 61. Головная боль (95)
- 62. Головная боль в височной части (80)
- 63. Головная боль в затылочной части (80)
- 64. Пульсирующая головная боль (80)
- 65. Сдавливающая головная боль (08)
- 66. Головокружение (60)
- 67. Обмороки или обморочное состояние (90)
- 68. Шум в ушах (65)
- 69. Шум в сердце (55)
- 70. Ухудшение зрения (65)
- 71. Видения перед глазами (95)
- 72. Потемнение в глазах (95)
- 73. Расширение зрачков (95)
- 74. Нарушение сознания (100)
- 75. Потеря ориентировки во времени и пространстве (90)
- 76. Заторможенность (75)
- 77. Рассеянность внимания (75)
- 78. Апатия (75)
- 79. Нарушение сна (95)
- 80. Нарушение речи (100)
- 81. Проблемы с глотанием (95)
- 82. Психомоторное возбуждение (75)
- 83. Шаткость походки (85)
- 84. Неустойчивость (90)
- 85. Ограничение двигательных функций (паралич или парез) (95)
- 86. Ограничение двигательных функций нижних конечностей (95)
- 87. Ограничение двигательных функций верхних конечностей (95)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

88. Ограничение двигательных функций нижних конечностей с одной стороны тела (95)
89. Ограничение двигательных функций верхних конечностей с одной стороны тела (95)
90. Чувство жара (65)
91. Озноб (70)
92. Зябкость пальцев (65)
93. Зябкость нижних конечностей (65)
94. Зябкость пальцев ног (65)
95. Зябкость верхних конечностей (65)
96. Зябкость пальцев рук (65)
97. Повышение температуры (95)
98. Низкая температура кожи (65)
99. Низкая температура пальцев (65)
100. Низкая температура ног (65)
101. Низкая температура пальцев ног (65)
102. Низкая температура рук (65)
103. Низкая температура пальцев рук (65)
104. Повышение давления (95)
105. Понижение давления (95)
106. Частые перепады настроения (55)
107. Покраснение кожи (90)
108. Покраснение лица (90)
109. Покраснение пальцев (90)
110. Покраснение нижних конечностей (90)
111. Покраснение пальцев ног (90)
112. Покраснение верхних конечностей (90)
113. Покраснение пальцев рук (90)
114. Покраснение ушей (90)
115. Покраснение губ (75)
116. Покраснение носа (90)
117. Покраснение глаз (95)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

- 118. Покраснение кожи головы (70)
- 119. Бледность кожи (75)
- 120. Бледность лица (75)
- 121. Бледность пальцев (75)
- 122. Бледность нижних конечностей (75)
- 123. Бледность пальцев ног (75)
- 124. Бледность верхних конечностей (75)
- 125. Бледность пальцев рук (75)
- 126. Бледность ушей (75)
- 127. Бледность губ (90)
- 128. Бледность носа (75)
- 129. Синюшность кожи (80)
- 130. Цианоз нижних конечностей (90)
- 131. Цианоз пальцев (90)
- 132. Цианоз пальцев ног (90)
- 133. Цианоз верхних конечностей (90)
- 134. Цианоз пальцев рук (90)
- 135. Цианоз лица (90)
- 136. Цианоз ушей (90)
- 137. Цианоз губ (80)
- 138. Цианоз носа (90)
- 139. Желтизна кожи (75)
- 140. Желтизна лица (75)
- 141. Сухость кожи (70)
- 142. Сухость кожи лица (70)
- 143. Сухость кожи ног (70)
- 144. Сухость кожи рук (70)
- 145. Влажность кожи (70)
- 146. Носовые кровотечения (100)
- 147. Внутренние кровотечения (50)
- 148. Набор веса (75)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

- 149. Потеря веса (75)
- 150. Дефицит веса (75)
- 151. Потеря аппетита (55)
- 152. Кашель (95)
- 153. Кашель усиливается в положении лёжа (80)
- 154. Сухой кашель (80)
- 155. Мокрота (90)
- 156. Хрипы (85)
- 157. Кровохарканье (95)
- 158. Набухание вен на шее (95)
- 159. Увеличение печени (60)
- 160. Боль в правом подреберье (90)
- 161. Вздутие живота (85)
- 162. Появление внутренних язв (55)
- 163. Появление язв на пальцах (100)
- 164. Появление язв на нижних конечностях (100)
- 165. Появление язв на пальцах ног (100)
- 166. Появление язв на верхних конечностях (100)
- 167. Появление язв на пальцах рук (100)
- 168. Судороги в нижних конечностях (95)
- 169. Судороги в нижних конечностях беспокоят по ночам (80)
- 170. Судороги в верхних конечностях (95)
- 171. Быстрая утомляемость ног (55)
- 172. Быстрая утомляемость рук (60)
- 173. Тяжесть в ногах (50)
- 174. Жжение в ногах (75)
- 175. Зуд кожи (65)
- 176. Зуд кожи лица (65)
- 177. Зуд кожи рук (65)
- 178. Зуд кожи ног (65)
- 179. Сосудистые звёздочки (95)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и



- 180. Сосудистые звёздочки на коже (95)
- 181. Сосудистые звёздочки на ногах (95)
- 182. Сосудистые звёздочки на руках (95)
- 183. Ассиметричная улыбка (90)
- 184. Потемнение кожи (90)
- 185. Потемнение нижних конечностей (95)
- 186. Потемнение пальцев (95)
- 187. Потемнение пальцев ног (95)
- 188. Потемнение верхних конечностей (95)
- 189. Потемнение пальцев рук (95)
- 190. Потемнение лица (95)
- 191. Потемнение ушей (95)
- 192. Потемнение губ (80)
- 193. Потемнение носа (95)
- 194. Ломкость волос (70)
- 195. Выпадение волос (80)
- 196. Выпадение волос на руках (80)
- 197. Выпадение волос на ногах (80)
- 198. Жажда (80)
- 199. Икота (90)
- 200. Олигурия (75)
- 201. Тремор (80)
- 202. Тремор нижних конечностей (80)
- 203. Тремор верхних конечностей (80)
- 204. Тремор пальцев (80)
- 205. Тремор пальцев ног (80)
- 206. Тремор пальцев рук (80)
- 207. Сыпь (85)
- 208. Сыпь на руках (85)
- 209. Сыпь на ногах (85)
- 210. Проблемы с памятью (65)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

- 211. "Болячки" на голове (65)
- 212. Импотенция (80)
- 213. Затруднённое мочеиспускание (95)
- 214. Гиперчувствительность кожи (70)
- 215. Гиперчувствительность пальцев (70)
- 216. Гиперчувствительность кожи нижних конечностей (70)
- 217. Гиперчувствительность пальцев ног (70)
- 218. Гиперчувствительность кожи верхних конечностей (70)
- 219. Гиперчувствительность пальцев рук (70)
- 220. Белок в моче (80)
- 221. Кровь в моче (80)
- 222. Диарея (85)
- 223. Жидкий стул (85)
- 224. Сонливость (65)
- 225. Аммиачный запах изо рта (80)
- 226. Гипергликемия (55)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

## ПРИЛОЖЕНИЕ 3

**ПЕРЕЧЕНЬ СЕРДЕЧНО\_СОСУДИСТЫХ ЗАБОЛЕВАНИЙ, ВЕРОЯТНОСТЬ  
НАЛИЧИЯ КОТОРЫХ МОЖЕТ БЫТЬ ОПРЕДЕЛЕНА СИМПТОМЧЕКЕРОМ**

1. Инфекционный эндокардит
2. Узелковый периартериит
3. Болезнь Такаясу
4. Болезнь Хортона
5. Атеросклероз почечной артерии
6. Острая почечная недостаточность
7. Кардиосклероз
8. Аритмия
9. Артериальная гипертензия
10. Парасимпатический криз при нейроциркуляторной дистонии
11. Гипотоническая нейроциркуляторная дистония
12. Гипертоническая нейроциркуляторная дистония
13. Симпатический криз при нейроциркуляторной дистонии
14. Кардиальная нейроциркуляторная дистония
15. Миокардиодистрофия
16. Ревмокардит
17. Миокардит
18. Тромбоэмболия лёгочной артерии
19. "Лёгочное сердце"
20. Стенокардия

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

- 21. Ишемическая болезнь сердца**
- 22. Инфаркт миокарда**
- 23. Проплапс митрального клапана**
- 24. Кардиомиопатия**
- 25. Перикардит**
- 26. Приобретённый порок сердца**
- 27. Врождённый порок сердца**
- 28. Правожелудочковая сердечная недостаточность**
- 29. Лёвожелудочковая сердечная недостаточность**
- 30. Атеросклероз артерий головного мозга**
- 31. Острое нарушение мозгового кровообращения**
- 32. Сосудистый коллапс**
- 33. Атеросклероз артерий нижних конечностей**
- 34. Варикозное расширение вен нижних конечностей**
- 35. Облитерирующий эндартериит**
- 36. Синдром Лериша**
- 37. Периферическая ангиопатия**
- 38. Облитерирующий тромбангиит**
- 39. Атеросклероз артерий верхних конечностей**
- 40. Синдром Рейно**

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

### КОД ПРОГРАММЫ, ГЕНЕРИРУЮЩЕЙ ДАТАСЕТ, НА КОТОРОМ ПРОИЗВОДИТСЯ ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ

```

import numpy as np

import random

# Ввод данных о порядковых номерах симптомов, характерных для заболевания,

# частоты встречаемости каждого характерного симптома при наличии заболевания,

# вероятностях корректности обнаружения каждого из характерных симптомов пользователем.

input_data = np.array (list(map(int, input().split())) - 1

occurencies = np.array (list(map(int, input().split()))

expected_reliabilities = np.array (list(map(int, input().split())) / 100

# Нахождение числа строк датасета, которые необходимо заполнить ненулевыми значениями для
каждого симптома.

# Нахождение числа случаев, когда у человека встретились все симптомы, характерные для
заболевания.

symptoms_count = len (input_data)

probability_of_abcent = -occurencies + 100

real_maximum = 1000 - np.prod (probability_of_abcent / 100) * 1000

cs_strings_count = max (1, int ((np.prod (occurencies / 100)) * 1000 / real_maximum * 1000))

occurencies = occurencies * 10 / real_maximum * 1000

occurencies = occurencies.astype(int) - cs_strings_count

remain_strings_count = 1000 - cs_strings_count

# Составление наборов случайных вещественных чисел, которыми необходимо заполнить датасет.

```

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

```
expected_reliabilities_table = [[0 for t in range (occurencies[i])] for i in range (len (input_data))]
```

```
for i in range (len (input_data)):
```

```
    for t in range (occurencies[i]):
```

```
        expected_reliabilities_table[i][t] = random.uniform(expected_reliabilities[i] - min
(expected_reliabilities[i] - 0.5, 1 - expected_reliabilities[i]), expected_reliabilities[i] + min
(expected_reliabilities[i] - 0.5, 1 - expected_reliabilities[i]))
```

```
    random.shuffle (expected_reliabilities_table[i])
```

# Заполнение датасета (data\_table) согласно приведённому в основной части данного документа алгоритму.

```
data_table = [[0 for i in range (227)] for i in range (remain_strings_count)]
```

```
for t in range (len (input_data)):
```

```
    for i in range (len (data_table)):
```

```
        data_table[i][0] = np.sum (np.array (data_table[i])) - data_table[i][0]
```

```
    data_table.sort()
```

```
    for k in range (occurencies[t]):
```

```
        data_table[k][input_data[t] + 1] = expected_reliabilities_table[t][k]
```

# Добавление в датасет случаев, когда у человека встретились все симптомы, характерные для заболевания.

```
add_data_table = [[0 for i in range (227)] for i in range (cs_strings_count)]
```

```
for i in range (cs_strings_count):
```

```
    for j in range (len (occurencies)):
```

```
        add_data_table[i][input_data[j] + 1] = random.uniform(expected_reliabilities[j] - min
(expected_reliabilities[j] - 0.5, 1 - expected_reliabilities[j]), expected_reliabilities[j] + min
(expected_reliabilities[j] - 0.5, 1 - expected_reliabilities[j]))
```

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

```
data_table += add_data_table
```

```
# Загрузка данных в файл.
```

```
file = open("dataset.txt", "a")
```

```
for i in range (len (data_table)):
```

```
    file.write (" ".join(map(str, data_table[i][1:]))+'\n')
```

```
file.close()
```

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.11-01 81 01-1				
Инв. № подл.	Подп. и дата	Взам. инв.	Инв. №	Подп. и

[illegible]



