

Автоматизация распознавания заявок радиослушателей

Д.М. Коробкин, С.А. Фоменков, Н. Брызгалин, А.А. Васильев

Волгоградский государственный технический университет

Аннотация: В статье представлено описание автоматизации процесса распознавания аудиозаписи с целью выявления заказанной песни на радиостанции. Использована модель распознавания русской речи Golos от SberDevices. Разработан алгоритм коррекции текста, полученного в результате анализа аудиозаписи с помощью модели Golos, на основе метода расстояния Левенштейна. Для распознанных заявок радиослушателей организовано взаимодействие с БД DIGISPOT II (формирование и выполнение запросов для поиска исполнителей и их песен).

Ключевые слова: распознавание речи, Golos, Digispot II.

Введение

В настоящее время взаимодействие радиодиджеев филиала ВГТРК «ГТРК «Волгоград-ТРВ» со слушателями радиоэфиров в режиме «Стола заказов» не автоматизировано, осуществляется запись звонка радиослушателя на автоответчик с последующим «ручным» прослушиванием заказа радиодиджеем. При этом могут возникать сложности в случае заказа песни, отсутствующей в ротации (базе аудиозаписей) радиоканала. Поскольку обратной связи в режиме «автоответчика» нет, то диджей в таком случае ставит любую песню на свой выбор.

Текущий процесс требует от диджея прослушивания всех накопившихся в течении недели звонков радиослушателей, причем длительность звонка варьируется от пары до нескольких десятков минут, причем не всегда непосредственно заказ песни следует в конце заявки. Несомненно, требуется автоматизация процесса распознавания речи и ключевых слов [1 – 3] с целью выявления заказанной песни.

Процесс обработки заявок начинается с приема входящего звонка от радиослушателя и начала его записи. Радиодиджей может записать поздравление без песни или с песней в подарок, а также заказать поставить определенную песню в эфир без поздравления. Далее происходит

распознавание речи в записанном файле разговора [4, 5], после чего в полученном тексте начинается поиск названия песни и имя автора, которые находятся в заполненной базе данных (БД) DIGISPOT II – встроенной системе автоматизации радиовещания в компании ВГТРК. Полученные результаты поиска, содержащие аудиофайл, текст распознанной речи, имя автора при наличии и название песни, отправляются радиодиджею.

Автоматизированный процесс представлен на рис. 1. Рассмотрим его подробнее.

Модуль работы со звонками использует GSM-модем [6] для приема звонков. С помощью него можно принимать звонки, СМС, а также инициировать исходящие вызовы. Как только модуль получает сообщение о входящем вызове, он принимается и начинается запись разговора. По сообщению об окончании разговора, считывание байт-записи прекращается, формируется массив-байт, содержащий аудиофайл и передается в модуль распознавания речи.

Модуль распознавания речи принимает файл записи разговора и возвращает текст разговора. Полученный текст передается в модуль поиска автора и песни. Список авторов и песен содержится в предварительно заполненной базе данных. Если найденный автор не является исполнителем найденной песни, происходит обращение к модулю передачи инструкций и обработки вызовов с целью инициации исходящего звонка для уточнения автора и песни, которую заказал пользователь. После чего снова происходит поиск. Если соответствия снова не найдено, то аудиофайл, расшифровка звонка и имя автора с названием песни отправляются в модуль оповещения радиодиджея. В случае, если найдена была только песня, то в модуль отправляется название песни и список авторов, которые исполняют песни с таким же названием. Если не было найдено ни песни, ни автора, то отправляется сообщение о том, что заказ на песню отсутствует.

Модуль формирования отчета формирует HTML-файл с результатами поиска, содержащий: запись разговора, распознанный текст, если поиск был успешен, то содержит также исполнителя и название песни, а также путь до файла во внутренней системе ВГТРК.

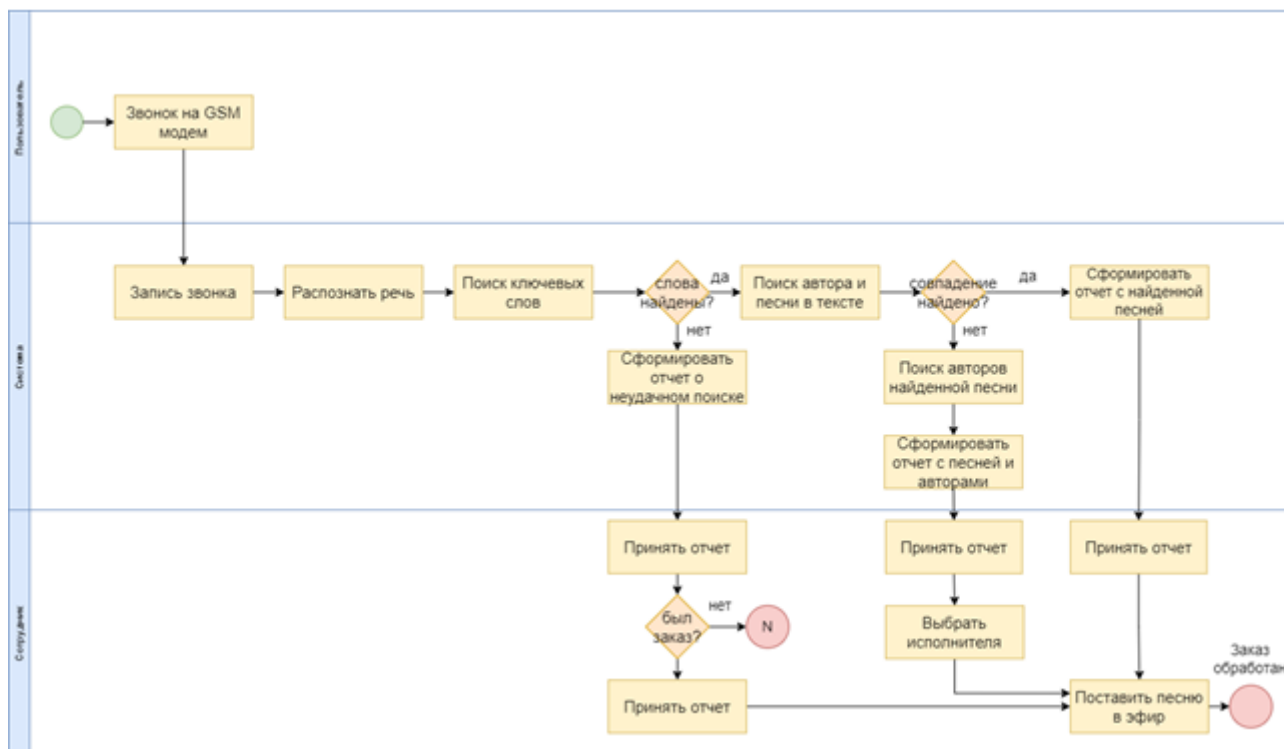


Рис. 1. – Описание автоматизированного процесса (Диаграмма TO-VE)

Модели и методы

В основе алгоритма поиска автора и песни в распознанном тексте используется метрика Левенштайна [7], позволяющая определить, какое количество операций вставки, замены и удаления необходимо провести, чтобы из одной строки получить другую. Это необходимо, поскольку в результате распознавания аудиозаписей с использованием модели Golos [8] формируется текст с большим количеством орфографических ошибок (анализ звукозаписи происходит на основе распознавания аллофон, т.е. фонетики, поэтому при плохом качестве записи или наличии дефектов речи происходит существенное искажение исходного произнесённого текста). Алгоритм представлен на рис. 2, 3.

Поиск происходит в два основных этапа. Первый этап заключается в поиске ключевых слов, таких как “песню”, “подарок” или ”исполнении”. Каждое слово ищется в исходном тексте и собираются индексы вхождения искомым слов. После чего выбирается последний индекс и текст обрезается до него.

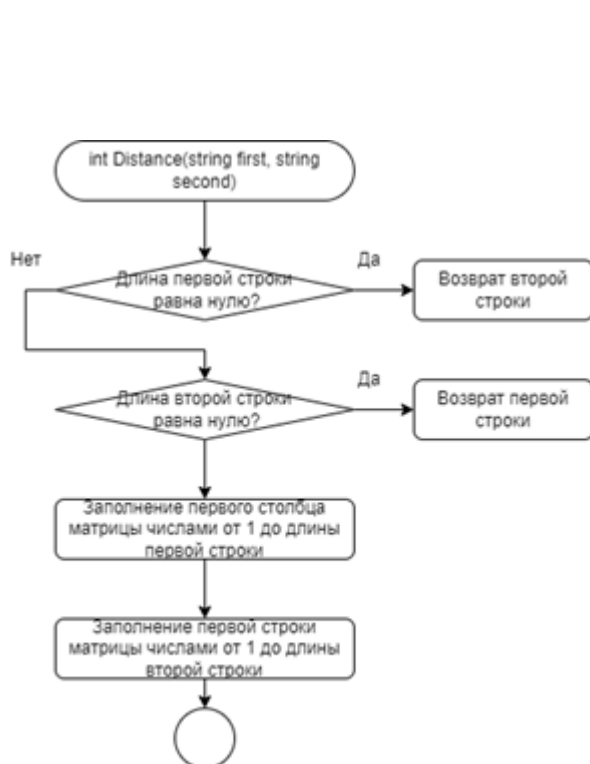


Рис. 2. – Проверка входных данных и предварительное заполнение матрицы

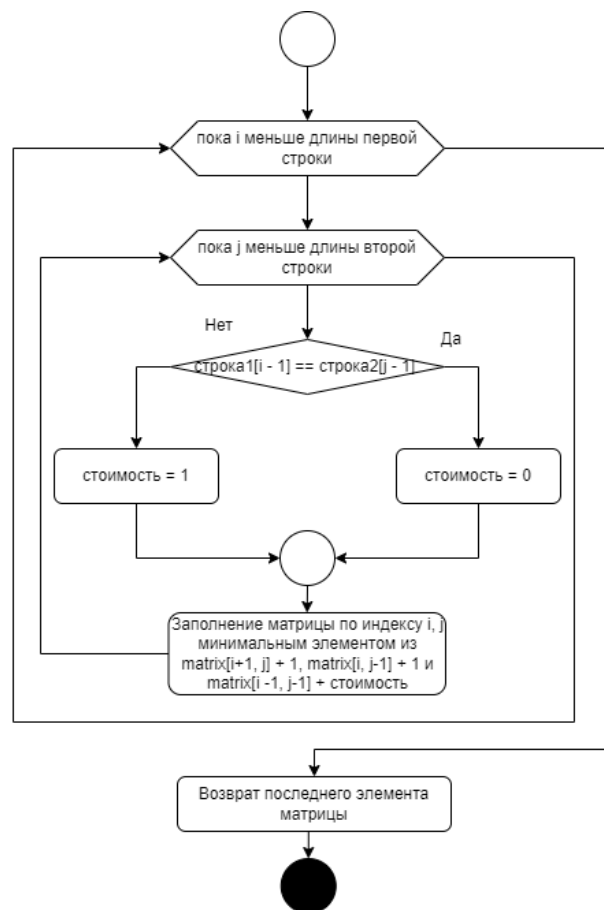


Рис. 3. – Вычисление расстояний Левенштейна

В основе первого этапа также лежит поиск расстояние Левенштейна. Ключевые слова выбраны в ходе анализа записей с автоответчика. Во всех случаях пользователь называет сначала слово “песню”, а потом ее название. Либо же “песню в подарок” или “песню в исполнении”, после чего обычно следует имя автора песни. Алгоритм представлен на рис. 4.

На втором этапе начинается поиск необходимых слов. Алгоритм функции представлен в приложении Б. Входной текст разделяется на отдельные слова, после чего запускается цикл с перебором всех строк из списка, содержащего все строки для поиска. В программе будут использоваться списки с авторами и названиями песен.

Каждая строка этого списка приводится к нормальной форме: убираются пробелы, запятые и другие знаки. В итоге получается последовательность из символов алфавита.

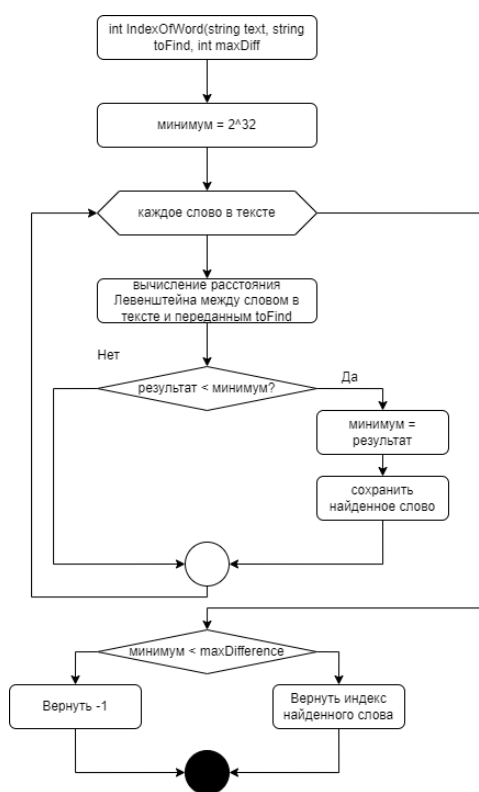


Рис. 4. – Алгоритм поиска слов в тексте

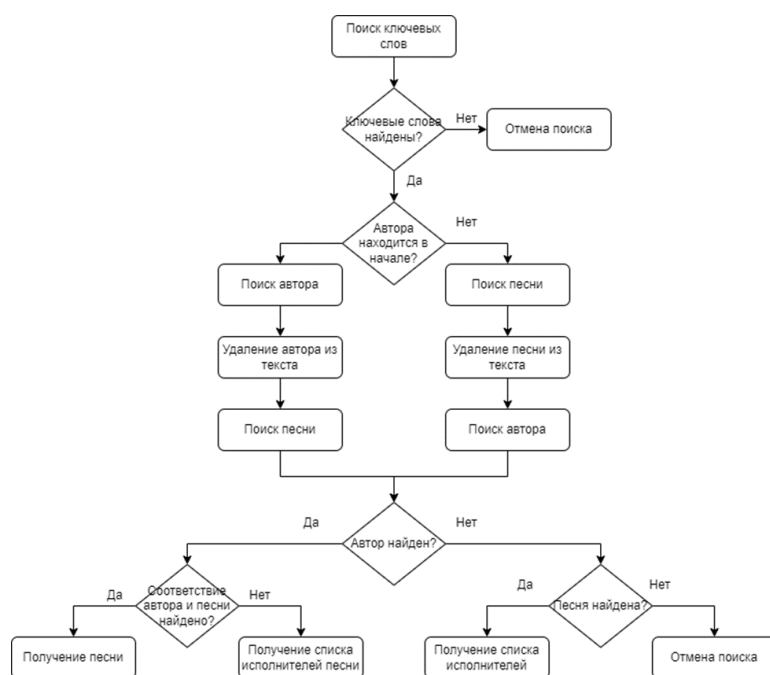


Рис. 5. – Диаграмма активности

Далее считается количество слов в каждой строке и определяется условие окончания следующего цикла. В цикле формируется строка из слов исходного текста, она и созданная заранее строка в нормальной форме отправляются в метод вычисления метрики Левенштейна.

Если расстояние меньше текущего, то обновляются минимум, найденное слово и найденное слово из списка. По окончании основного

цикла, найденное минимальное значение сравнивается с пороговым и если минимальное значение меньше порогового, то возвращаются две строки: найденная строка в тексте и соответствующая ей строка из списка. Иначе возвращаются пустые строки, означая, что поиск не дал результатов.

С использованием описанных выше реализованных функций, происходит поиск автора и песни. Поиск основан на шаблонах, которые были выявлены в ходе анализа записей с автоответчика. Шаблоны представлены на рис. 6.



Рис. 6. – Визуальное представление шаблонов заказов

Синим цветом отмечены ключевые слова, желтым - места, где находится название песни и красным - места, где находится имя автора.

Для того, чтобы учесть все шаблоны, в строке ищутся и вырезаются названия песни или имя автора, чтобы увеличить точность поиска. Полученные результаты по окончании поиска используются для формирования отчета в соответствующем модуле.

Диаграмма активности данного процесса представлена на рис. 5.

Проектирование программы

Программа должна обеспечивать возможность выполнения перечисленных ниже функций:

- прием записей звонков пользователей;
 - распознавание записанной речи;
 - поиск песен;
-

- поиск исполнителей;
- формирование отчета, содержащего информацию, какую песню какого исполнителя радиослушатель запросил поставить на радио.

Входные данные:

- Входящий звонок по сети сотовой связи и полученная с помощью нее запись разговора любой длительности в одноканальном режиме, частотой 8 кГц и передачей 16 бит в единицу времени.

Выходные данные:

- Папка с названием “call_<дата окончания звонка>”, содержащая запись разговора в формате wave и сформированным отчетом в формате HTML.

Архитектура программы представлена на рис. 7.



Рис. 7. – Архитектура программы

Модуль автоматизации приема и обработки заявок от пользователей реализован на языке C# 10 и целевой платформе .NET 6.x [9].

Подмодуль работы со звонками работает напрямую с GSM модемом. Работа модуля с модемом осуществляется через COM порты, посредством получения и отправки AT-команд [10]. Модем занимается двумя портами:

голосовым и командным. Через голосовой порт считывается поток байтов записанной речи.

Подмодуль распознавания речи использует акустическую модель GOLOS от SBER для распознавания речи. Она принимает на вход аудиофайлы формата .mp3 или .wav в формате моно. Запись звонка с GSM-модема возможна сразу в моноформате, из-за чего отпадает необходимость конвертации.

Для работы подмодуля поиска автора и песни используется заранее заполненная база данных, которая встроена в систему автоматизации радиовещания DIGISPOT II. Поиск основан на вычислении метрики Левенштейна. Для повышения шансов успешного поиска реализован алгоритм, который проверяет текст на соответствие шаблонам заказа, выявленных в ходе анализа существующих записей автоответчика.

Результаты

Пример распознанного из аудиозаписи текста:

*«Дорогая передача ее ведущей двадцать третьего января ученической
людмила павловной круглая дата людоочка возбавляю себя днем рождения
говорит добного здоровья семейного благополучия чтобы в вашей семье
было бы побольше вазе содеть взаимопонимания чтобы деть вновь и
правники помогали в трупную себе минуту маливиночка я благодарю тебя за
все хорошее оставляю такой же добрый отзывившай душой к людям
опоздравляет себя любиящая тебя мария о песню в подарок пименинница
спасибо всем ведущим этой передаче дохранит всех господь»*

Найденное в результате название песни: «Именинница»

Найденный автор: не распознан

Исполнители песен с таким же названием в БД DIGISPOT II (рис. 9):

- *Е.А. Верховская*

- *Елена Василёк (Елена Верховская)*

Распознанный текст	Найденная песня	Автор	Путь до файла
каки дорогая передача ее ведущей двадцать третьего января круглая дата любочка возбавляю себя днем рождения говорит добного здоровья семейного благополучия чтобы в вашей семье было бы побольше вазе содеть взаимопонимания чтобы деть вновь и правники помогли в трупную себе минуту маливиночка я благодарю тебя за все хорошее оставляя такой же добрый отзывившей душой к людям опоздравляет себя любящая тебя мария о песню в подарок пимениница спасибо всем ведущим этой передаче дохранит всех господь	Именинница	Песни для эфира	\\Vog.radio.gtrk\DFS\TRACT\ROOT\SND\MUSIC\Лена Василёк - Именинница (audiopoisk.com).mp3_(551E311D).wav
		Елена Василёк (Елена Верховская)	\\Vog.radio.gtrk\DFS\TRACT\ROOT\SND\MUSIC\Лена Василёк - Именинница (audiopoisk.com).mp3_(551E311D).wav
		Е.А. Верховская, сл Е.А. Верховская (Семин В.В.)	\\Vog.radio.gtrk\DFS\TRACT\ROOT\SND\MUSIC\Лена Василёк - Именинница (audiopoisk.com).mp3_(551E311D).wav

Рис. 8. – Найденные песни в БД DIGISPOT II

Заключение

В результате проведенной работы была спроектирована и реализована программа для автоматизации приема и обработки заявок от радиослушателей.

Область применения разработанной системы – автоматизация режима «Стола заказов» на радио и взаимодействия радиодиджеев филиала ВГТРК «ГТРК «Волгоград-ТРВ»» со слушателями. Получившаяся модульная система позволяет заменить каждый из них на другой, так что существует возможность в дальнейшем заменить необходимые модули и использовать данную разработку на других радиостанциях.

Основным направлением совершенствования и дальнейшего развития разработанной системы является замена оборудования для приема звонков с целью повышения качества получаемой записи разговора. Благодаря такой замене, появится возможность более эффективно осуществлять поиск необходимых данных в распознанной речи.

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 24-21-20140, rscf.ru/project/24-21-20140/ и Администрации Волгоградской области.

Литература

1. Bin Sahbudin M. A., Chaouch C., Scarpa M., Serrano S. IoT based Song Recognition for FM Radio Station Broadcasting // 7th International Conference on Information and Communication Technology (ICoICT). Kuala Lumpur, Malaysia: IEEE, 2019. URL: researchgate.net/publication/335864365_IoT_based_Song_Recognition_for_FM_Radio_Station_Broadcasting.
2. Rose R.C., Paul D.B. A hidden Markov model based keyword recognition system // International Conference on Acoustics, Speech, and Signal Processing. Albuquerque, NM, USA: IEEE, 1990. URL: ieeexplore.ieee.org/document/115555.
3. Lileikytė R., Lamel L., Gauvain J., Gorin A. Conversational telephone speech recognition for Lithuanian // Computer Speech & Language. 2018. №49. URL: [sciencedirect.com/science/article/abs/pii/S0885230816300523](https://www.sciencedirect.com/science/article/abs/pii/S0885230816300523).
4. Волченков В.А. Обнаружение пауз в речевых сигналах // Инженерный вестник Дона. 2013. №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2024.
5. Астапов К.А. Применение вейвлет-преобразования для сокращения области значения искусственных нейронных сетей на примере задачи распознавания речи // Инженерный вестник Дона. 2009. №1. URL: ivdon.ru/ru/magazine/archive/n1y2009/105.
6. Асташов Д.А. Обзор GSM модулей // Электронный журнал: наука, техника и образование. 2019. №2. URL: elibrary.ru/item.asp?id=39175027.
7. Коробкин Д. М. Система автоматического субтитрования видеофайлов // Системный анализ в науке и образовании. 2022. №2. URL: sanse.ru/index.php/sanse/article/view/532.



8. Сбер открывает доступ к датасету Golos — самому большому набору речевых данных на русском, размеченному вручную // SberDevices URL: sberdevices.ru/press/detail/sber_otkryvaet_dostup_k_datasetu_golos_samomu_bolsomu_naboru_recevykh_dannykh_na_russkom_razmecennomu_vruchnuyu/ (дата обращения: 29.04.2024).

9. Hosting services in .NET Core console application // Think Rethink URL: pgroene.wordpress.com/2018/08/02/hostbuilder-ihost-ihostedservice-console-application/ (дата обращения: 29.04.2024).

10. AT Commands, GSM AT command set // Engineers Garage URL: engineersgarage.com/at-commands-gsm-at-command-set/ (дата обращения: 29.04.2024).

References

1. Bin Sahbudin M. A., Chaouch C., Scarpa M., Serrano S. 7th International Conference on Information and Communication Technology (ICoICT). Kuala Lumpur, Malaysia: IEEE, 2019. URL: researchgate.net/publication/335864365_IoT_based_Song_Recognition_for_FM_Radio_Station_Broadcasting.

2. Rose R.C., Paul D.B. International Conference on Acoustics, Speech, and Signal Processing. Albuquerque, NM, USA: IEEE, 1990. URL: ieeexplore.ieee.org/document/115555.

3. Lileikytė R., Lamel L., Gauvain J., Gorin A. Computer Speech & Language. 2018. №49. URL: [sciencedirect.com/science/article/abs/pii/S0885230816300523](https://www.sciencedirect.com/science/article/abs/pii/S0885230816300523).

4. Volchenkov V.A. Inzhenernyj vestnik Dona. 2013. №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2024.

5. Astapov K.A. Inzhenernyj vestnik Dona. 2009. №1. URL: ivdon.ru/ru/magazine/archive/n1y2009/105.

6. Astashov D.A. Elektronij zhurnal: nauka, tekhnika i obrazovanie. 2019. №2. URL: elibrary.ru/item.asp?id=39175027.



7. Korobkin D. M. Sistemnyj analiz v nauke i obrazovanii. 2022. №2. URL: sanse.ru/index.php/sanse/article/view/532.

8. Sber otkryvaet dostup k datasetu Golos - samomu bol'shому набору rechevyh dannyh na russkom, razmechennomu vruchnuju [Sber opens access to the Voice dataset, the largest set of manually labeled speech data in English]. SberDevices URL: sberdevices.ru/press/detail/sber_otkryvaet_dostup_k_datasetu_golos_samomu_bol_somu_naboru_recevyx_dannyx_na_russkom_razmecennomu_vruchnuju/ (data access: 29.04.2024).

9. Hosting services in .NET Core console application. Think Rethink URL: pgroene.wordpress.com/2018/08/02/hostbuilder-ihost-ihostedservice-console-application/ (data access: 29.04.2024).

10. AT Commands, GSM AT command set. Engineers Garage URL: engineersgarage.com/at-commands-gsm-at-command-set/ (data access: 29.04.2024).

Дата поступления: 10.05.2024

Дата публикации: 19.06.2024