

## Алгоритмическое обеспечение чат-бота для консультации потенциальных абитуриентов

*Д.Н. Куций*

*Южно-Российский государственный политехнический университет (НПИ)  
имени М.И. Платова, Новочеркасск*

**Аннотация:** В данной статье определяются основные принципы функционирования чат-бота. Выполнено моделирование последовательности процессов. Статья содержит описание алгоритма работы чат-бота. Для визуализации указанного алгоритма используются диаграммы деятельности и состояний. Представлена физическая модель хранения необходимых для проведения консультации данных. Рассмотрены вопросы нечеткого поиска по ключевым словам. В статье приведено описание процесса исправления ошибок с использованием метрики  $n$ -граммных расстояний. Продемонстрировано функционирование чат-бота.

**Ключевые слова:** чат-бот, диалоговая система, мессенджер, обработка естественного языка, нечеткий поиск, расстояние Дамерау-Левенштейна,  $n$ -грамма.

### Введение

Современный уровень развития мобильных устройств обуславливает повышение спроса на удобные и эффективные способы обмена информацией и взаимодействия с различными сервисами и приложениями. Популярными инструментами в современной коммуникации и предоставлении услуг в различных сферах деятельности, включая образование, бизнес, клиентское обслуживание и многое другое [1-2], стали чат-боты – программы-собеседники, в основе которых лежит некоторый сценарий общения с пользователем [3].

В продвижении бренда университета и повышении информированности абитуриентов о вузе также могут быть задействованы автоматизированные виртуальные консультанты [4].

Целью работы является повышение эффективности взаимодействия абитуриентов с приемной комиссией учебного заведения.

Для достижения указанной цели необходимо получить комплексное решение [5], в котором будут учтены последовательность процессов и

условия, при которых будут совершены операции, а также обработка и хранение данных.

### Функциональное моделирование

С учетом постановки задачи построено системное представление программы-собеседника. Формализация выполнена с помощью методологии комплексного описания Integrated Definition (IDEF) уровня функционального моделирования IDEF0. На рис. 1 представлена общая схема использования чат-бота.



Рис. 1. – Контекстная диаграмма основного процесса

Основная функция, выполняемая системой, – консультирование потенциальных абитуриентов. Входным потоком данных этого процесса является запрос пользователя. Конечный итог выполняемой функции – предоставленная абитуриенту информация.

В качестве управляющего воздействия используется информация из базы данных, а механизмом, выполняющим действия в системе, является программное обеспечение бота.

Декомпозиция контекстной диаграммы изображена на рис. 2. Управляющее воздействие «Правила обработки конструкций естественного языка» и результат «Команда отсутствует» помечены квадратными скобками,

так как они были добавлены для детализации первого уровня и отсутствовали на родительской диаграмме.

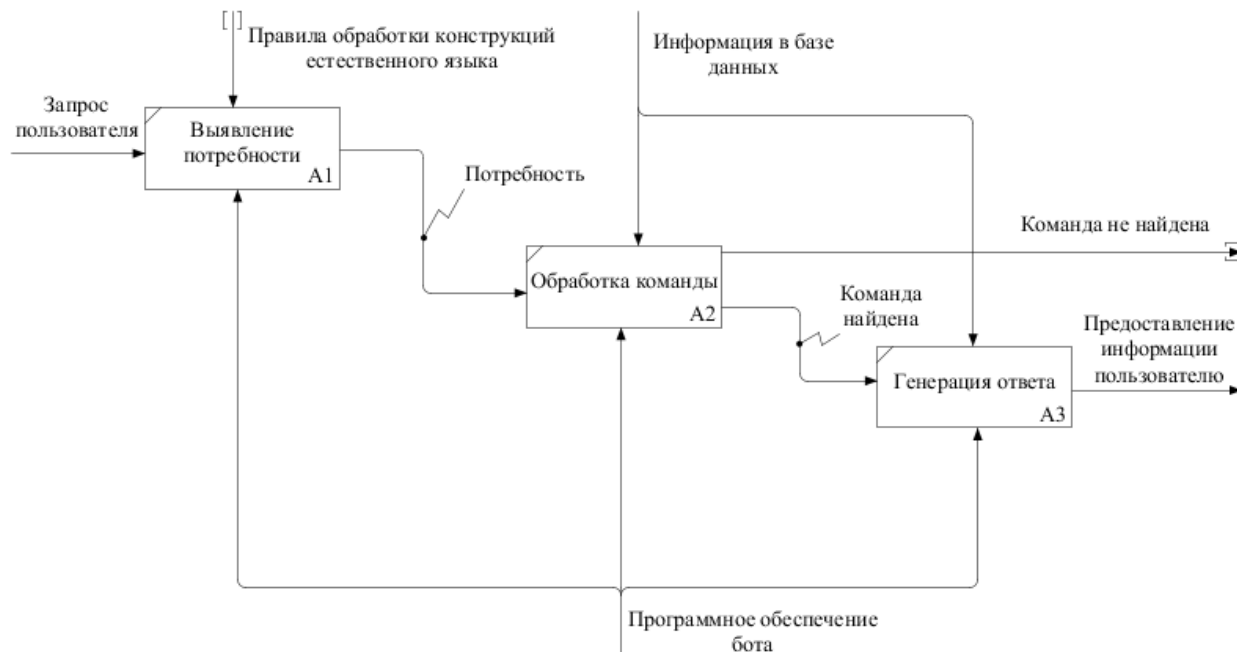


Рис. 2. – Декомпозиция контекстной диаграммы

Графическая интерпретация логики взаимодействия элементов системы, последовательность процессов и их условность (рис. 3) выполнена также с помощью методологии IDEF на уровне документирования IDEF3 [6].



Рис. 3. – Моделирование последовательности процессов

Логический узел  $J_1$  соответствует поддерживаемому набору команд  $X$ , а  $J_2$  отображает разветвление процессов в зависимости от результата сопоставления сообщения пользователя с существующими шаблонами команд.

## Общий алгоритм работы

Пользователь отправляет запрос боту, запрос поступает на сервер, где происходит анализ и поиск нужного ответа. При нахождении подходящего ответа, бот выводит ответ на экран. При необходимости, диалог продолжается. Если запрос не может быть обработан, пользователю выдается сообщение об ошибке. Визуализация описанного алгоритма представлена на рис. 4, в виде диаграммы деятельности.

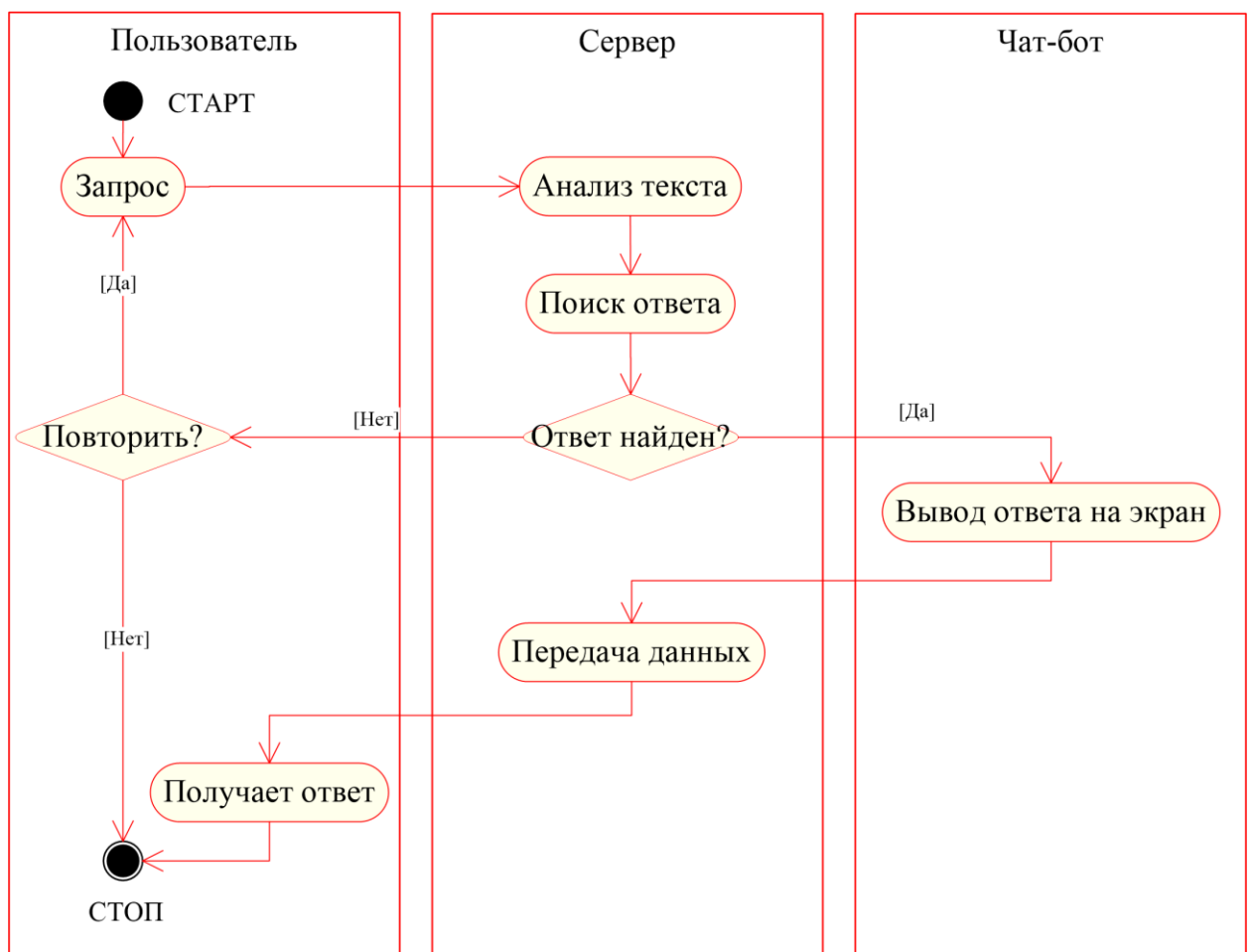


Рис. 4. – Диаграмма деятельности

На рис. 5 представлена диаграмма состояний, которая рассматривает работу чат-бота и его реакции на действия пользователя, ищущего необходимую информацию.

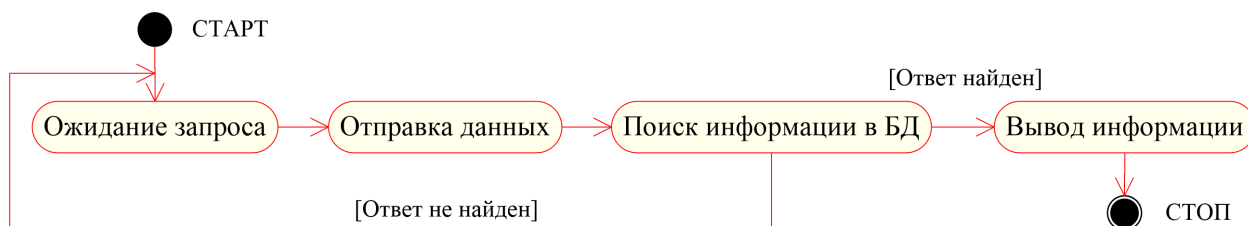


Рис. 5. – Диаграмма состояний

### Модель хранения данных

На рис. 6 представлена физическая модель данных на уровне объектов системы управления базой данных.

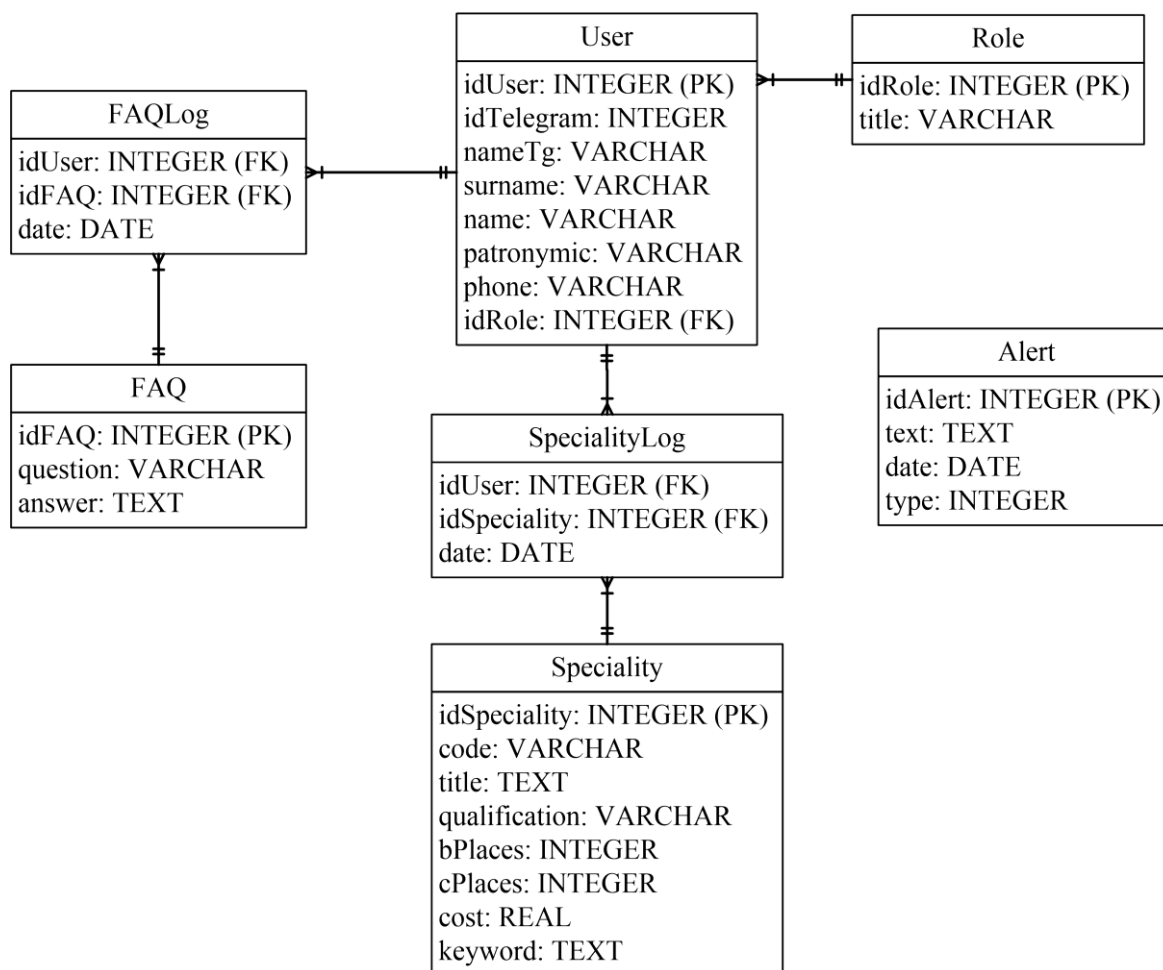


Рис. 6. – Физическая модель данных

Таблица «User» содержит информацию о пользователях чат-бота, «Role» – разграничивает функционал пользователей. В «Speciality» хранится информация о направлениях подготовки, «FAQ» – список наиболее

задаваемых вопросов. «SpecialityLog» и «FAQLog» нужны для формирования статистики обращений абитуриентов.

Таблица Alert используется для хранения актуальных объявлений о процессе приемной кампании. Она не связана с другими таблицами, так как информация в ней доступна всем пользователям.

### **Обработка сообщений пользователя**

В основе ведения беседы чат-ботом лежит поиск по ключевым словам или словосочетаниям в хранящемся в базе данных списке часто задаваемых вопросов [7], с учетом частичного совпадения и наличия ошибок или опечаток (рис. 7).

Решение задачи устранения ошибки в запросе, отправляемом абитуриентом программе, может заключаться в создании списка возможных слов [8].

Эффективным подходом является использование  $n$ -граммного разложения слов, при котором слова разбиваются на подстроки определенной длины  $n$  с последующим вычислением функции сходства на основе общих фрагментов [9].

Процесс исправления ошибок с помощью метрики  $n$ -граммных расстояний включает в себя следующую последовательность действий:

1. В зависимости от длины слова, выбирается способ его разбиения, и затем создается список биграмм или триграмм.
2. Для каждой  $n$ -граммы формируется рейтинг терминов в виде инвертированного индекса.
3. Выполняется обработка компонентов разложения запроса и хранимой в базе данных информации.
4. Для экономии времени при обработке будущих запросов формируется словарь терминов с ошибками и список кандидатов с их рейтингами.

С целью уменьшения количества возможных ответов, выдаваемых абитуриенту, используется второй уровень обработки результатов с помощью расстояния редактирования.

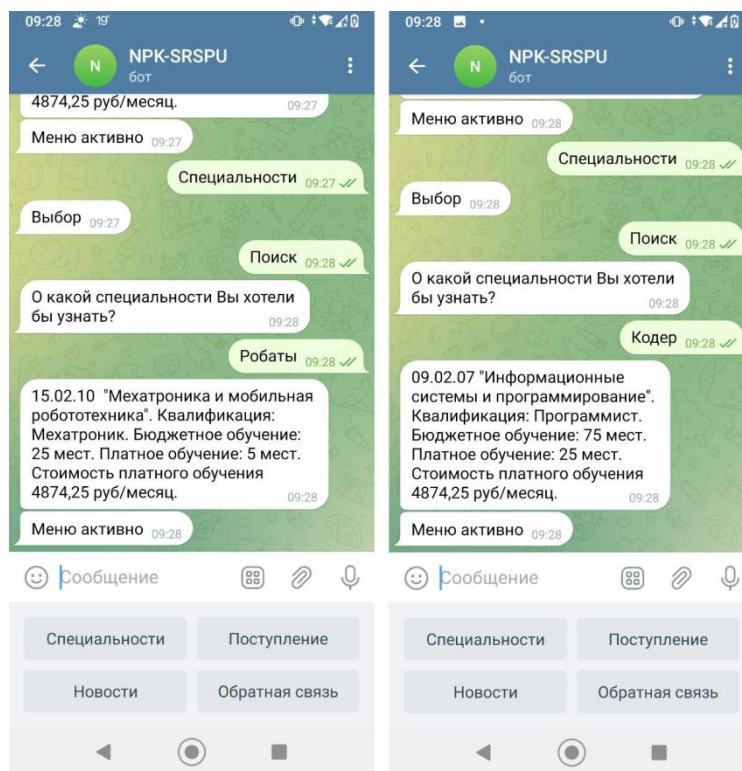


Рис. 7. – Результаты поиска информации по ключевым словам

В рассматриваемой реализации используется коэффициент, вычисляемый на основе метрики расстояния Дамерау-Левенштейна. Критическое значение максимального расстояния между словами – количества операций редактирования – устанавливается пропорционально длине исходного запроса для обеспечения наиболее релевантных результатов.

Расстояние Дамерау-Левенштейна для двух строк  $S_1$  и  $S_2$  длиной  $l_1$  и  $l_2$ , соответственно, равно [10]:

$$d(S_1, S_2) = D(l_1, l_2) = \sum_{i=0, j=0}^{l_1, l_2} D_{ij} = \sum_{i=0, j=0}^{l_1, l_2} \min(X_{\text{в}}, X_{\text{у}}, X_{\text{з}} + C_{\text{замены}}, X_{\text{т}} + C_{\text{транспозиции}}),$$

где – метрика для символов  $S_1[i]$  и  $S_2[j]$ ;  $X_{\text{в}}$  – цена вставки символа  $S_2[j]$ ;  $X_{\text{у}}$  – цена удаления символа  $S_1[i]$ ;  $X_{\text{з}}$  – цена замены символа  $S_1[i]$  на символ  $S_2[j]$ ;

$X_T$  – цена перестановки символа  $S_1[i]$  на символ  $S_2[j-1]$  или  $S_1[i-1]$  на символ  $S_2[j]$ .

Поправочные коэффициенты замены и транспозиции вычисляются по формулам:  $C_{\text{замены}} = \begin{cases} 1, S_1[i] \neq S_2[j] \\ 0, \text{иначе} \end{cases}$  и  $C_{\text{транспозиции}} = \begin{cases} 1, S_1[i] = S_2[j-1], S_1[i-1] = S_2[j] \\ 0, \text{иначе} \end{cases}$ .

В итоговой матрице редактирования для нахождения не более  $k$  различий, нужно вычислять лишь диагональную полосу шириной  $2k+1$ , что сводит временную сложность к  $O(k \min(l_1, l_2))$ .

### Заключение

Рассмотренное в статье алгоритмическое обеспечение комплексного решения на базе мобильных технологий, диалоговых систем и нечеткого поиска, может рассматриваться в качестве инструмента повышения эффективности взаимодействия абитуриентов с приемной комиссией учебного заведения, а также улучшения качества информированности о вузе и продвижения бренда университета.

### Литература

1. Чеха А.М., Литвиненко Е.Ю., Куций Д.Н. Разработка и применение чат-бота в школьном образовании // Наука. Образование. Культура. Вклад молодых исследователей: VI Междунар. науч. конф. препод., молодых ученых, аспирантов и студентов вузов, г. Новочеркасск, 26 апр. 2022 г. / Юж.-Рос. гос. политехн. ун-т (НПИ). Новочеркасск: Лик, 2022. С. 236-239.
2. Маслова М. А., Бажутова Д. А., Дмитриев А. С. Алгоритмы работы чат-бота для поиска товаров // Инженерный вестник Дона, 2021, №4. URL: [ivdon.ru/ru/magazine/archive/n4y2021/6921](http://ivdon.ru/ru/magazine/archive/n4y2021/6921).
3. Misischia C., Poczte F., Strauss C. Chatbots in customer service: Their relevance and impact on service quality // Procedia Computer Science. 2022. Vol.201. pp. 421-428. URL: <https://doi.org/10.1016/j.procs.2022.03.055>.





4. Суханова Н.Т., Вежелис Т.М. Использование чат-ботов для автоматизации предоставления справочной информации абитуриентам и студентам вузов // Проблемы современного педагогического образования. 2022. №76-2. С. 178-181.

5. Ashfaq M., Yun J., Yu S., Loureiro S. I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents // Telematics and Informatics. 2020. Vol.54. С. 101473. URL: doi.org/10.1016/j.tele.2020.101473.

6. Куликова Я.В., Качалов Д.Л. Метод определения эмоционального состояния человека при помощи чат-бота // Инженерный вестник Дона, 2022, №9. URL: ivdon.ru/ru/magazine/archive/n9y2022/7893/.

7. Козицын А.С., Афонин С.А., Шачнев Д.А. Алгоритм поиска по ключевым словам специалистов в заданной предметной области // Современные информационные технологии и ИТ-образование. 2021. №1, т.17. С. 124-133.

8. Белов С.Д., Зрелова Д.П., Зрелов П.В., Кореньков В.В. Обзор методов автоматической обработки текстов на естественном языке // Системный анализ в науке и образовании. 2021. №3. С. 1-15.

9. Гринченков Д.В., Куций Д.Н. Решение задачи построения запросов в системе тематического поиска на основе распознавания частично структурированных текстов // Известия высших учебных заведений. Северо-Кавказский регион. Технические науки. 2019. №1. С. 10-16.

10. Гринченков Д.В., Куций Д.Н. Актуальность и принципы построения интеллектуальной информационной системы формирования методического обеспечения учебных дисциплин на основе ресурсов сети Интернет // Известия высших учебных заведений. Северо-Кавказский регион. Серия: Технические науки. 2014. №3. С. 114-119.

## References

1. Chekha A.M., Litvinenko E.Yu., Kushchiy D.N. Nauka. Obrazovanie. Kul'tura. Vklad molodykh issledovateley: VI Mezhdunar. nauch. konf. prepod., molodykh uchenykh, aspirantov i studentov vuzov, g. Novocherkassk, 26 apr. 2022. Yuzh.-Ros. gos. politekhn. un-t (NPI). Novocherkassk: Lik, 2022. pp. 236-239.
2. Maslova M.A., Bazhutova D.A., Dmitriev A.S. Inzhenernyj vestnik Dona, 2021, №4. URL: [ivdon.ru/ru/magazine/archive/n4y2021/6921](http://ivdon.ru/ru/magazine/archive/n4y2021/6921).
3. Misischia C., Poetze F., Strauss C. Procedia Computer Science. 2022. Vol.201. pp. 421-428. URL: [doi.org/10.1016/j.procs.2022.03.055](https://doi.org/10.1016/j.procs.2022.03.055).
4. Sukhanova N.T., Vezhelis T.M. Problemy sovremennogo pedagogicheskogo obrazovaniya. 2022. №76-2. pp. 178-181.
5. Ashfaq M., Yun J., Yu S., Loureiro S. Telematics and Informatics. 2020. Vol.54. p. 101473. URL: [doi.org/10.1016/j.tele.2020.101473](https://doi.org/10.1016/j.tele.2020.101473).
6. Kulikova Ya.V., Kachalov D.L. Inzhenernyj vestnik Dona, 2022, №9. URL: [ivdon.ru/ru/magazine/archive/n9y2022/7893/](http://ivdon.ru/ru/magazine/archive/n9y2022/7893/).
7. Kozitsyn A.S., Afonin S.A., Shachnev D.A. Sovremennye informatsionnye tekhnologii i IT-obrazovanie. 2021. №1, t.17. pp. 124-133.
8. Belov S.D., Zrelva D.P., Zrelv P.V., Koren'kov V.V. System analysis in science and education. 2021. №3. pp. 1-15.
9. Grinchenkov D.V., Kushchiy D.N. Bulletin of Higher Educational Institutions. North Caucasian Region. Technical Sciences. 2019. №1. pp. 10-16.
10. Grinchenkov D.V., Kushchiy D.N. Bulletin of Higher Educational Institutions. North Caucasian Region. Technical Sciences. 2014. №3. pp. 114-119.