

УДК 004.89

DOI: [10.26102/2310-6018/2022.37.2.014](https://doi.org/10.26102/2310-6018/2022.37.2.014)

## Чат-бот с использованием технологий нейронных сетей и методов обработки текста для повышения лояльности клиентов

Д.А. Сюсюра<sup>1</sup>✉, А.В. Коваленко<sup>1</sup>, М.В. Шарпан<sup>2</sup>

<sup>1</sup>Кубанский государственный университет, Краснодар, Российская Федерация

<sup>2</sup>Краснодарский университет МВД России, Краснодар, Российская Федерация  
[Darick2137@yandex.ru](mailto:Darick2137@yandex.ru)✉

**Резюме.** В эпоху цифровизации в обществе основным инструментом социального взаимодействия человека становятся мобильные устройства. А с ростом популярности мессенджеров место чат-ботов в мобильной среде становится все значительнее. Интеллектуальные интерактивные чат-боты часто используются в мобильных приложениях и способствуют улучшению взаимодействия между компаниями и их заказчиками, что в итоге повышает лояльность клиентов этой организации. Чат-боты позволяют компаниям общаться с заказчиками в индивидуальном порядке, не привлекая к этой работе сотрудников и тем самым экономя временной, денежный и человеческий ресурс. Большинство чат-ботов работают по алгоритмам сценарного подхода и не являются универсальными. Это обусловлено простотой и скоростью разработки. Однако в таком случае существует риск упустить множество вариантов в дереве решений. Эту проблему способны решить чат-боты, основанные на нейронных сетях, но стоит учитывать, что и те и другие имеют недостаток в виде долгой обработки сообщений и обратной связи. В случае со сценарным подходом это происходит из-за долгих перемещений по веткам. Для нейронных сетей возникает сложность из-за алгоритма обработки обратной связи. В таком случае использование сервиса не будет оправдано, лояльность клиентов к организации будет падать. В связи с чем в статье рассматривается альтернативный подход к созданию чат-ботов на основе технологий нейронных сетей и методов представления текста, который позволяет избежать описанных выше проблем. В качестве технологий для реализации чат-бота были использованы: Python 3.6, библиотеки genism, sklearn, scipy, pandas, технология word2vec и doc2vec. Также в статье описан способ ускорения получения обратной связи и обучения чат-бота с помощью K-мерных деревьев.

**Ключевые слова:** нейронные сети, чат-бот, технология word2vec, мессенджеры, обработка текста.

**Для цитирования:** Сюсюра Д.А., Коваленко А.В., Шарпан М.В. Чат-бот с использованием технологий нейронных сетей и методов обработки текста для повышения лояльности клиентов. *Моделирование, оптимизация и информационные технологии.* 2022;10(2). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1110> DOI: 10.26102/2310-6018/2022.37.2.014

## Chatbot based on neural networks and word embedding to increase customer loyalty

D.A. Syusyura<sup>1</sup>✉, A.V. Kovalenko<sup>1</sup>, M.V. Sharpan<sup>2</sup>

<sup>1</sup>Kuban State University, Krasnodar, Russian Federation

<sup>2</sup>Krasnodar University of the Ministry of Internal Affairs of Russia, Krasnodar, Russian Federation  
[Darick2137@yandex.ru](mailto:Darick2137@yandex.ru)✉

**Abstract.** In the digital era mobile devices are becoming the main instrument of human social interaction. With the growing popularity of instant messengers, the role of chatbots in the mobile environment appears to be more and more significant. Intelligent interactive chatbots are often used in mobile applications and help improve the interaction between companies and their customers, which ultimately increases customer loyalty to that organization. Chatbots allow companies to communicate with customers on an individual basis, without involving employees and thereby saving time, money, and human resources. The majority of chatbots works with scripted algorithms and they are not universal. This is due to the simplicity and speed of development. However, in this case, there is a risk of missing many choices in the decision tree. Chatbots based on neural networks can solve this problem, but it should be taken into consideration that both of them have a drawback – long processing of messages and feedback. In the context of the scenario approach, this is caused by long branch transitions. For neural networks, complexity arises because of the feedback processing algorithm. In that instance, the application of the service will not be justified, customer loyalty to the organization will deteriorate. In this connection, the article discusses an alternative approach to creating chatbots with the aid of neural network technologies and text representation methods, which avoids the problems described above. As a means of chatbot design, the following technologies were utilized: Python 3.6, genism libraries, sklearn, scipy, pandas, word2vec and doc2vec technology. The article also describes a way to accelerate chatbot feedback and training using KD-Trees.

**Keywords:** neural networks, chat bot, word2vec technology, messengers, word embedding.

**For citation:** Syusyura D.A., Kovalenko A.V., Sharpan M.V. Chatbot based on neural networks and word embedding to increase customer loyalty. *Modeling, Optimization and Information Technology*. 2022;10(2). Available from: <https://moitvvt.ru/ru/journal/pdf?id=1110> DOI: 10.26102/2310-6018/2022.37.2.014 (In Russ.).

## Введение

Предметом исследования в данной статье является чат-бот на основе нейронных сетей с оптимизированной структурой данных на основе KD-Tree для ускоренной обработки обратной связи от пользователя.

Чат-боты уже давно стали неотъемлемой частью ведения современного бизнеса, поскольку позволяют автоматизировать типовые процессы, экономить человеческие ресурсы и деньги компании. [1-2] Однако примитивные чат-боты, основанные на сценарном подходе, чаще вредят, поскольку усложняют процесс поиска информации для клиента. Человеку приходится пройти по длинной сценарной ветке, прежде чем он наконец получит ответ на свой вопрос. В связи с чем возникает актуальность нейросетевого подхода в проектировании чат-ботов. Самый эффективный подход к проектированию строится на основе векторного представления текста и технологий word2vec и doc2vec.

Существует множество чат-ботов, каждый из которых обеспечивает поддержку определенной сферы [3]. Однако универсального решения нет, либо оно дорогостоящее, поскольку требует переучивания модели нейронной сети, что довольно небыстрый и энергозатратный процесс.

В статье представлена реализация чат-бота «Bot Bot», не требующего переобучения или дообучения под конкретную сферу после того, как модель нейронной сети была обучена в первый раз. Для его реализации была выбрана архитектура чат-бота, основанная на нейронных сетях с применением технологии word2vec и doc2vec, а также возможностью получения обратной связи от пользователя. Последний критерий особенно важен, так как позволяет чат-боту самообучаться.

Целью исследования является рассмотрение преимуществ чат-ботов на основе нейронных сетей с технологией word2vec в сравнении со сценарным подходом и другими NLP методами, а также решение проблемы оптимизации времени при обработке запросов пользователей. На основе указанных преимуществ предложена реализация чат-

бота, способного быть адаптивным под разные сферы деятельности без переобучения за счет технологии word2vec с ускоренной обработкой запросов.

Можно обозначить 3 задачи исследования:

- 1) Выделение преимуществ чат-ботов на основе нейронных сетей.
- 2) Общие принципы работы и практическое применение алгоритма word2vec для создания чат-бота.
- 3) Сокращение сложности и времени обработки запросов для созданного чат-бота за счет оптимизации структуры данных на основе KD-Tree.

### Материалы и методы

Чат-бот должен уметь определять лучший ответ на любое полученное сообщение, понимать намерения отправителя и тип ответного сообщения, формировать грамматически и лексически правильный отклик. Современные чат-боты сталкиваются с трудностями при решении этих задач. Для преодоления этих трудностей используются нейронные сети с глубоким обучением [4]. В этом случае чат-бот использует некоторые варианты модели от последовательности к последовательности (Seq2Seq).

В основе word2vec и doc2vec лежат технологии Skip-Gram и CBOW (Continuous Bag of Words). Оба алгоритма используют близлежащие слова для извлечения семантики слов. В Skip-Gram предсказываются слова контекста, используя главное слово. В Continuous Bag of Words алгоритм выполняет противоположную операцию: из контекстных слов модель предсказывает главное слово. Согласно оригинальной статье [5] обнаружено, что Skip-Gram хорошо работает с небольшими наборами данных и может лучше представлять редко встречающиеся слова. Однако CBOW обучается быстрее, чем Skip-Gram, и может лучше представлять часто встречающиеся слова. Конечно, выбор модели во многом зависит от задачи, которая ставится. В нашем случае важно представлять редкие слова, поэтому выбор был сделан в пользу Skip-Gram.

Векторное представление текста [6] позволяет представить отображение слова из естественного языка в вектор фиксированной длины (обычно в пределах от 1 до 5 сотен измерений, чем выше это значение, тем точнее представление, но сложнее его вычислить). В некотором семантическом пространстве каждое слово переводится в точку, причем операциям над смыслом слов соответствуют алгебраические операции. Рассмотрим архитектуру нейронной сети, представленной на Рисунке 1.

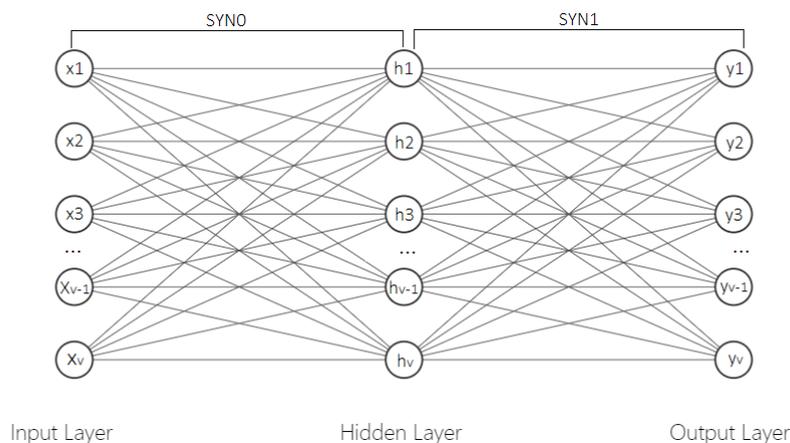


Рисунок 1 – Архитектура нейронной сети  
 Figure 1 – Neural network architecture

Для реализации Word2vec [6-7] используется один скрытый слой. Размерность входного слоя равна количеству слов в словаре. Размер скрытого слоя соответствует размерности пространства. Размер выходного идентичен входному. Таким образом, если словарь для обучения состоит из  $V$  слов и  $N$  размерности векторов слов, веса между входным и скрытым слоем образуют матрицу  $SYN0$  размера  $V \times N$ . Которая выглядит на Рисунке 2 следующим образом:

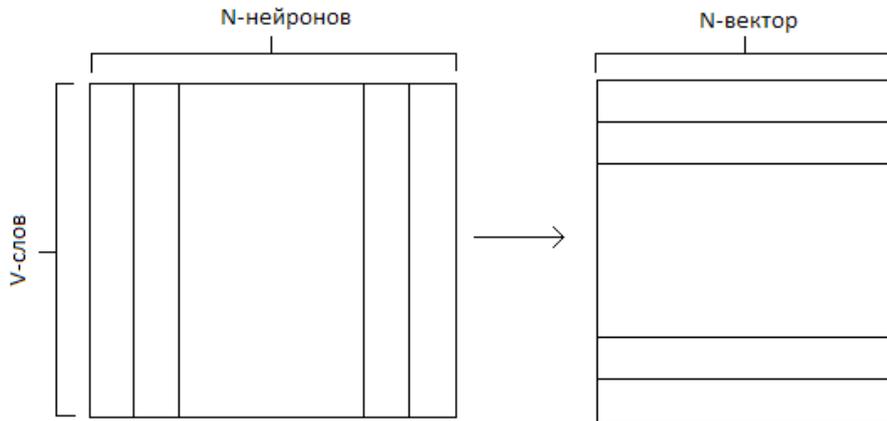


Рисунок 2 – Матрица  $SYN0$   
 Figure 2 – Matrix  $SYN0$

Каждая из  $V$  строк является векторным  $N$ -мерным представлением слова (1):

$$h = SYN0^T x \tag{1}$$

Аналогично, веса между скрытым и выходным слоем образуют матрицу  $SYN1$  размера  $N \times V$ . Тогда на входе выходного слоя будет (2):

$$val_i = SYN1_j^T h \tag{2}$$

где  $SYN1_j$  –  $j$ -ый столбец матрицы  $SYN1$ .

Как было описано выше, выходной слой содержит  $N$  нейронов с функцией активацией softmax. Каждый нейрон соединен со скрытым слоем (слоем, где и происходит words embedding). Рисунок 3 показывает эти соединения:

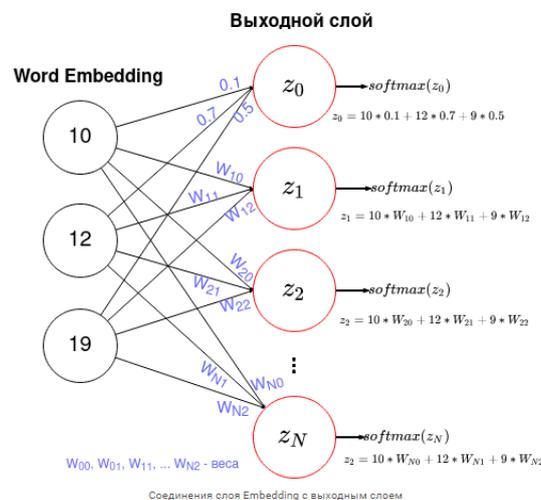


Рисунок 3 – Соединение скрытого слоя, где происходит векторизация текста  
 Figure 3 –Hidden layer connection where a word is embedded

Скалярное произведение – косинус угла между двумя точками в n-мерном пространстве. Эта формула позволяет определить, как близко находятся векторы слов. Если слова противоположные по значению, то произведение принимает значение -1. Затем используем «функцию мягкого максимума» (softmax), чтобы получить распределение слов.

Для реализации чат-бота сначала построим 300-мерное семантическое пространство на основе русскоязычной Википедии. Создадим экземпляр класса на Python для последующего обучения с параметрами минимальной частоты; промежутков, в котором рассматривается контекст; размерность вектора; максимальная частота появления слов; количество потоков. Затем строится таблица словарей.

В построенном семантическом пространстве, используя базу данных вопросов и ответов, найдем координаты всех фраз. Таким образом, не придется переучивать систему с расширением базы данных, достаточно учитывать добавленные фразы и находить их координаты в том же пространстве. Быстрая адаптация к обновлению данных – это основное достоинство разработанного чат-бота [8].

Схема работы чат-бота показана на Рисунке 4:

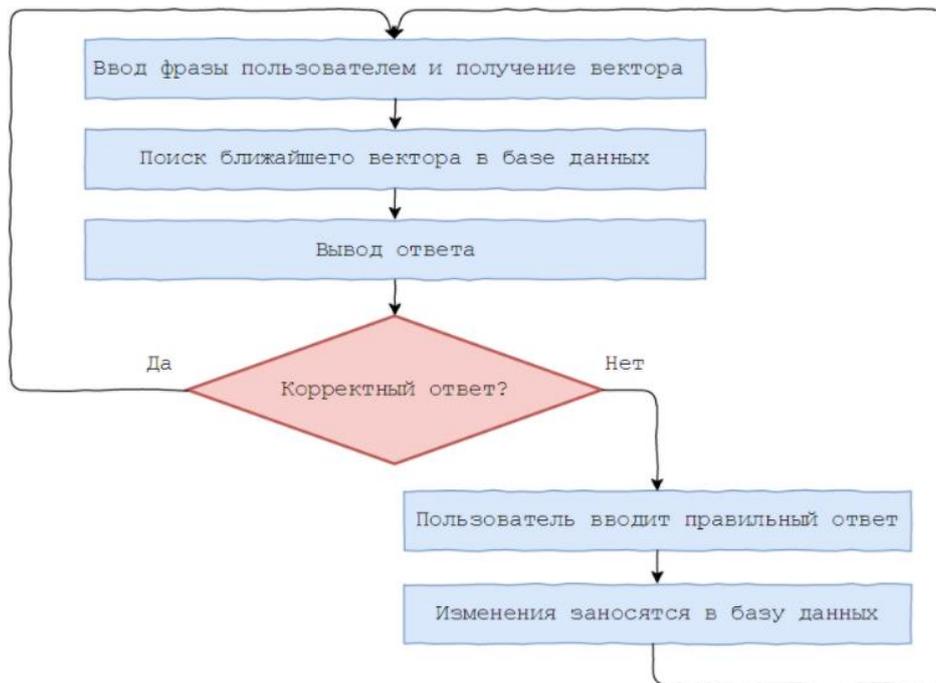


Рисунок 4 – Схема работы чат-бота  
 Figure 4 – Chatbot operation scheme

Для возможности обработки обратной связи от пользователя был использован метод построения KD-Tree. KD-Tree (K-мерное дерево) – структура данных, которая позволяет разбить K-мерное пространство на пространства меньшей размерности посредством отсечения гиперплоскостями [9-10]. Но сложность перестроения дерева при таком алгоритме составляет в среднем  $O(N \log N)$ . Решением для оптимизации данного алгоритма является – перестройка дерева каждые M добавлений фраз в базу данных, благодаря чему поиск происходит за  $O(\log N)$ .

Для обучения чат-бота был введен следующий функционал: после ответа на вопрос пользователю отображаются две кнопки для оценки качества ответа. В случае

отрицательного ответа пользователю предлагается скорректировать его, и исправленный результат заносится в базу данных.

### Результаты

В результате исследования был реализован прототип чат-бота на основе нейронных сетей и векторного представления текста с оптимизированным алгоритмом обработки обратной связи от пользователя за счет перестроения  $K$ -мерного дерева каждые  $M$  добавлений фраз в базу данных. В качестве технологий для его создания были использованы: Python 3.6, библиотеки `genism`, `sklearn`, `scipy`, `pandas`, технология `word2vec` и `doc2vec`. Практическая часть работы представлена прототипом AI-чат-бота на платформе мессенджера Telegram, способного адаптироваться под изменение базы данных и определенную сферу работы. Прототип чат-бота реализован на основе нейронной сети, реализующей концепцию внимания и определяющей намерение пользователя. На Рисунке 5 представлен прототип созданного чат-бота, а также пример взаимодействия с ним:

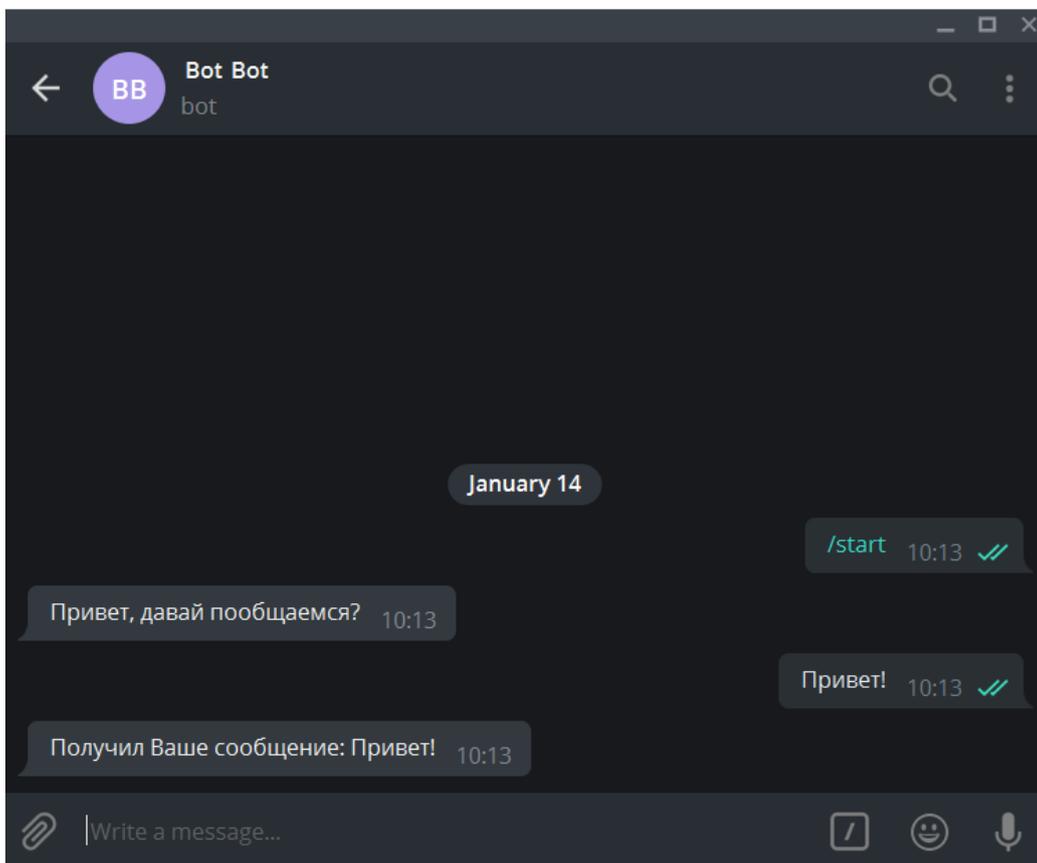


Рисунок 5 – Прототип чат-бота  
 Figure 5 – Chatbot prototype

Предложенный прототип AI-чат-бота реализован с помощью алгоритмов обмена сообщениями с пользователями через мессенджер Telegram, машинного обучения, преобразования данных в требуемый формат, а также работы с базой данных. Был усовершенствован алгоритм обработки обратной связи для сокращения времени добавления откорректированных ответов в базу данных. Пример получения обратной связи для обучения представлен на Рисунке 6:

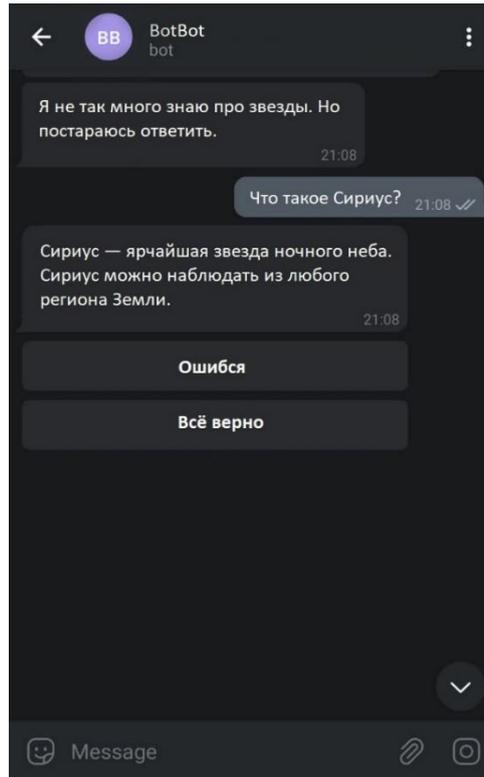


Рисунок 6 – Функционал получения обратной связи от пользователя  
Figure 6 – Functionality of user feedback

Поскольку текущая реализация чат-бота не способна распознавать ошибки в словах, был введен шаблонный ответ в виде текстового сообщения для запросов, классифицированных как «невозможные для распознавания». В случае получения слова с ошибкой или не существующего в словаре слова, чат-бот ответит сообщением «Я Вас не совсем понял!». Пример работы с такими запросами представлен на Рисунке 7.

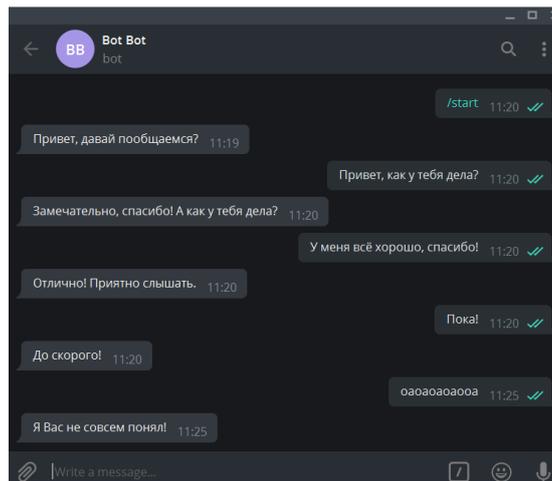


Рисунок 7 – Пример обработки некорректного запроса  
Figure 7 – An example of invalid request processing

### Обсуждение

Таким образом, разработанный прототип чат-бота на основе нейронных сетей имеет ряд преимуществ по сравнению со стандартным подходом к решению этой задачи

– с помощью сценария. Он адаптивен к любой сфере применения за счет того, что не требует постоянного переобучения. Его достаточно обучить 1 раз и затем лишь дополнять базу данных недостающей информацией с помощью функционала получения обратной связи. Кроме того, за счет оптимизации метода KD-Tree удалось сократить время обработки запроса пользователя. Эти ключевые моменты позволят снизить риски потери лояльности клиентов, а кроме того, денежные и временные потери на разработку, сделав ее более универсальной и адаптивной.

### Заключение

В результате проведенной работы был создан прототип чат-бота «Bot Bot», не требующего переобучения или дообучения под конкретную сферу после того, как модель нейронной сети была обучена в первый раз. Для его реализации была выбрана архитектура чат-бота, основанная на нейронных сетях с применением технологии word2vec и doc2vec, принципы работы которых разбирались в статье. Также в статье рассмотрена возможность получения обратной связи от пользователя и способ ее оптимизации с помощью перестроения KD-Tree через определенное количество итераций. Последний критерий особенно важен, так как позволяет чат-боту быстро самообучаться, за счет чего он существенно превосходит примитивные чат-боты, а также некоторые чат-боты, основанные на классическом использовании нейронных сетей.

### СПИСОК ИСТОЧНИКОВ

1. Chakrabarti C., Luger G.F. Artificial conversations for customer service chatter bots: architecture, algorithms, and evaluation metrics. *Expert Syst. Appl.* 2015;2(20):6878–6897.
2. Xu A., Liu Z., Guo Y., Sinha V., Akkiraju R. A new chatbot for customer service on social media. *Proceedings of CHI*. 2017:3506–3510.
3. Robert Dale. The Return of the Chatbots. *Natural Language Engineering*. 2016;22(5):811–817. DOI:10.1017/S1351324916000243.
4. Близнюк Б.О., Васильева Л.В., Стрельников И.Д., Ткачук Д.С. Современные методы обработки естественного языка. *Вісник Харківського національного університету імені В. Н. Каразіна*. 2017;36:14–26.
5. Mikolov T., Chen K., Corrado G., Jeffrey D. *Efficient estimation of word representations in vector space*. 2013; CoRR,abs/1301.3781.
6. Иванова А.В., Филиппова Л.Б. Методы обработки текста и машинного обучения при создании чат-ботов. *Новые горизонты. VIII научно-практическая конференция с международным участием. Сборник материалов и докладов*. 2021:289–292.
7. Mikolov T., Sutskever I., Chen K., Corrado G.S. and Dean J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013*. 2013;3111–3119.
8. Долгошей А.В. Применение нейронных сетей для обработки текста *Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях. Материалы XXIII Республиканской научной конференции студентов и аспирантов*. 2020;42–43.
9. Fengquan Zhang, Yahui Gao, Liuqing Xu. An adaptive image feature matching method using mixed Vocabulary-KD tree Multimedia. *Tools and Applications volume 2020*. 2020;79:16421–16439.
10. Лебединская Н.А., Лебединский Д.М. Обобщенные KD-деревья и локальные преобразования. *Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия*. 2012;3:42–43.

## REFERENCES

1. Chakrabarti C., Luger G.F. Artificial conversations for customer service chatter bots: architecture, algorithms, and evaluation metrics. *Expert Syst. Appl.* 2015;2(20):6878–6897.
2. Xu A., Liu Z., Guo Y., Sinha V., Akkiraju R. A new chatbot for customer service on social media. *Proceedings of CHI.* 2017:3506–3510.
3. Robert Dale. The Return of the Chatbots. *Natural Language Engineering.* 2016;22(5):811–817. DOI:10.1017/S1351324916000243.
4. Bliznuk B.O., Vasilieva L.V., Strelnikov I.D., Tkachuk D. S. Modern methods of natural language processing. *Bulletin of the Kharkov National University named by V.N. Karazin.* 2017;36:14–26.
5. Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean. Efficient estimation of word representations in vector space. 2013; CoRR,abs/1301.3781
6. Ivanova A.V., Filippova L.B. Word processing and machine learning techniques for creating chat bots. *In the collection: New horizons. VIII scientific and practical conference with international participation. Collection of materials and reports.* 2021; 289-292. (In Russ.)
7. Mikolov T., Sutskever I., Chen K., Corrado G.S. and Dean J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013.* 2013;3111–3119.
8. Dolgoshey A.V. Application of neural networks for text processing. *New mathematical methods and computer technologies in design, production and scientific research. Materials of the XXIII Republican Scientific Conference of Students and Postgraduates.* 2020;42–43. (In Russ.)
9. Fengquan Zhang, Yahui Gao, Liuqing Xu. An adaptive image feature matching method using mixed Vocabulary-KD tree Multimedia. *Tools and Applications volume 2020.* 2020;79:16421–16439.
10. Lebedinskaya N.A., Lebedinsky D.M. Generalized KD-trees and local transformations. *Bulletin of St. Petersburg University. Maths. Mechanics. Astronomy.* 2012;3:42–43. (In Russ.)

## ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Сюсюра Дарья Алексеевна**, магистр Кубанского государственного университета, Краснодар, Российская Федерация.  
email: [Darick2137@yandex.ru](mailto:Darick2137@yandex.ru)

**Syusyura Daria Alekseevna**, Master's Student, Kuban State University, Krasnodar, Russian Federation.

**Коваленко Анна Владимировна**, доктор технических наук, доцент, зав. каф. анализа данных и искусственного интеллекта Кубанского государственного университета, Краснодар, Российская Федерация.  
email: [savanna-05@mail.ru](mailto:savanna-05@mail.ru)

**Kovalenko Anna Vladimirovna**, Doctor of Technical Sciences, Associate Professor, Head of Department of Data Analysis and Artificial Intelligence of Kuban State University, Krasnodar, Russian Federation.

**Шарпан Мария Владимировна**, кандидат физико-математических наук, старший преподаватель Краснодарского университета МВД России, Краснодар, Российская Федерация.  
email: [marusi2000@mail.ru](mailto:marusi2000@mail.ru)

**Sharpan Maria Vladimirovna**, Candidate of Physical and Mathematical Sciences, Senior Lecturer of Krasnodar University of the Ministry of Internal Affairs of Russia, Krasnodar, Russian Federation.

*Статья поступила в редакцию 12.12.2021; одобрена после рецензирования 29.04.2022;  
принята к публикации 25.05.2022.*

*The article was submitted 12.12.2021; approved after reviewing 29.04.2022;  
accepted for publication 25.05.2022.*