

УДК 004.912

DOI: [10.26102/2310-6018/2024.45.2.042](https://doi.org/10.26102/2310-6018/2024.45.2.042)

Применение диалога на естественном языке в дистанционном обучении

В.В. Братищенко✉

Байкальский государственный университет, Иркутск, Российская Федерация

Резюме. Актуальность исследования обусловлена низким уровнем применения диалога на естественном языке в дистанционном обучении. Создание таких средств на основе искусственного интеллекта делает процесс дистанционного обучения более доступным и привлекательным. В статье предлагается строить диалог на основе эталонных вопросов к содержанию дистанционного курса. Ответ выбирается на основе близости вопроса пользователя к эталону. В качестве множества ответов рекомендуется использовать структурные единицы дистанционного курса, а в качестве эталонных вопросов – соответствующие заголовки. Данные учебного диалога запоминаются и используются для расширения списка эталонных вопросов и обучения системы. Для контроля усвоения используется мера близости ответов обучаемого на контрольные вопросы и правильных вариантов ответов. Для формирования контрольных вопросов можно использовать словари дистанционных курсов и тестовые задания. Определение меры близости двух текстов предлагается выполнять с использованием косинуса эмбедингов наиболее близких термов. Данные сравнения текстов по предложенной методике подтверждают ее возможность правильно оценивать близость текстов и обосновывают ее применение для организации диалога на естественном языке в дистанционном обучении.

Ключевые слова: дистанционное обучение, ранжирующий чат-бот, диалог на естественном языке, эмбединг, мягкое тестирование, мера близости предложений.

Для цитирования: Братищенко В.В. Применение диалога на естественном языке в дистанционном обучении. *Моделирование, оптимизация и информационные технологии.* 2024;12(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1589> DOI: 10.26102/2310-6018/2024.45.2.042

Application of natural language dialogue in distance learning

V.V. Bratischenko✉

Baikal State University, Irkutsk, the Russian Federation

Abstract: The relevance of the study is due to the low level of use of dialogue in natural language in distance learning. The creation of such tools based on artificial intelligence will make the process of distance learning more accessible and attractive. The article proposes to build a dialogue based on standard questions for the content of the distance learning course. The answer is selected based on the similarity of the user's question to the standard. It is recommended to use the structural units of the distance learning course as a set of answers, and the corresponding headings as standard questions. The training dialogue data is remembered and used to expand the list of standard questions and train the system. To control learning, a measure of the similarity of the student's answers to test questions and the correct answer options is used. To generate test questions, you can use distance learning dictionaries and test tasks. It is proposed to determine the measure of similarity of two texts using the cosine of the embeddings of the closest terms. Data from comparing texts using the proposed methodology confirm its ability to correctly assess the similarity of texts and justify its use for organizing dialogue in natural language in distance learning.

Keywords: distance learning, ranking chatbot, natural language dialogue, embedding, soft testing, sentence similarity measure.

For citation: Bratishchenko V.V. Application of natural language dialogue in distance learning. 2024;12(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1589> DOI: 10.26102/2310-6018/2024.45.2.042 (In Russ.).

Введение

Взаимодействие пользователей с компьютерными программами смещается в сторону применения естественного языка. Первыми от запросов с применением математической логики отказались поисковые системы, сегодня они перешли на запросы на естественном языке с последующим разбором и формализацией потребностей. Жесткий диалог в виде заранее определенного набора команд не привлекает пользователей и «машинным» стилем, и необходимостью изучения этих команд. Эта тенденция еще сильнее проявляется в такой области как обучение, в которой общение является самым привлекательным способом передачи знаний.

В компьютерном обучении все еще преобладает жесткий стиль общения. Учебный материал преподносится в структуре страниц, сгруппированных в разделы, подразделы и параграфы. Применение словарей и гиперссылок создает определенную гибкость, но не меняет существенно predetermined систему просмотра. Компьютерные тесты реализуют еще более жесткую систему проверки знаний – каждое задание предусматривает определенное множество правильных ответов.

Одним из основных недостатков дистанционного обучения является недостаточный уровень общения с обучаемым [1]. Общение в чатах по расписанию противоречит самой сути асинхронного обучения. Поэтому актуальной является задача замены преподавателя искусственным интеллектом для имитации общения на темы, связанные с процессом обучения. Для устранения перечисленных недостатков предлагается применение чат-ботов [2–4].

В работе [2] на основе анкетирования выявлен значительный интерес обучающихся к диалогу с компьютером и определены основные функции образовательных чат-ботов: персонализация коммуникаций, рекомендации учебного контента, ответы на типичные вопросы, определение целей обучения и отслеживание прогресса.

Достаточно подробный список функций образовательного чат-бота представлен в работах [3, 4]:

- администрирование и управление – чат-боты используются для описания организации учебной деятельности;
- ответы на часто задаваемые вопросы – чат-бот обеспечивает обратную связь по часто задаваемым вопросам по административным или образовательным темам;
- наставничество – чат-бот используется для когнитивного и эмоционального мониторингов результатов обучения учащихся;
- мотивация – чат-бот обеспечивает эмоциональную и мотивационную поддержку обучающихся;
- тренировка конкретных навыков и способностей – чат-бот используется в качестве помощника по практике, такие тренажеры применяются при изучении языков, методов коммуникации, программирования и т. д.;
- рефлексия и метакогнитивные стратегии – чат-бот используется в качестве умелого собеседника, который помогает в обучении;
- оценка успеваемости учащихся – чат-бот оперативно измеряет результаты обучения.

По результатам обзора [5] в этот список можно добавить ссылки к справочным и методическим материалам, диагностику пробелов в знаниях и выделение учебных

элементов для дополнительного изучения. В том же обзоре выявлено, что учащиеся, в первую очередь, получают пользу от чат-ботов в трех ключевых областях: помощь в выполнении домашних заданий, персонализированное обучение, опыт обучения и развития различных навыков. Для педагогов главными преимуществами являются:

- экономия времени – преподаватели экономят время на рутинных задачах, включая составление расписания, выставление оценок и предоставление информации учащимся, что позволяет им уделять больше времени учебным занятиям, методической подготовке и вовлечению студентов;

- улучшенная педагогика – преподаватели могут использовать чат-боты с искусственным интеллектом, чтобы дополнить свои инструкции и обеспечить персонализированную поддержку.

В обзоре выделены следующие проблемы применения чат-ботов:

- надежность и точность, связанные с применением искусственного интеллекта (чат-боты с искусственным интеллектом могут давать предвзятые или неточные ответы);

- недостатки в формировании критического мышления;

- слабая вовлеченность в обучение и мотивация;

- недостаточная справедливость оценки.

Одна из проблем, с которой сталкиваются преподаватели при интеграции чат-ботов в образовании – сложность оценки работы учащихся, особенно когда дело доходит до письменных заданий или ответов. Преподавателям может быть сложно понять, являются ли ответы действительно созданными студентами или предоставлены искусственным интеллектом, что влияет на точность выставления оценок.

При рассмотрении технической составляющей чат-бот-приложений [6] выявляются три типа внутренней организации логики работы программ: чат-боты могут быть реализованы на наборе бизнес-правил, на искусственном интеллекте или объединять эти две логики в одной программе. Чат-боты с применением искусственного интеллекта выбирают ответы из списка заранее заготовленных ответов (ранжирующие чат-боты) или генерируют ответы с применением некоторой модели (генерирующие чат-боты).

Во всех цитируемых работах отмечается положительный эффект от замены предопределенного сценария чата диалогом на естественном языке.

На основании приведенного обзора можно сделать следующие выводы:

- 1) Общение на естественном языке повышает качество, эффективность и привлекательность дистанционного обучения.

- 2) Разработано достаточно много компьютерных средств обработки текстов на естественном языке, которые можно использовать в диалогах с обучаемыми.

- 3) Применение общения на естественном языке должно минимизировать риски неточного информирования обучаемых и повышать точность измерения учебных достижений.

Целью данного исследования является разработка методов применения диалога на естественном языке в дистанционном обучении. Для этого решаются следующие задачи:

- организация диалога для ответов на вопросы обучающихся по содержанию дистанционного курса;

- организация диалога для проведения контрольного опроса и оценивание близости ответа обучаемого правильным вариантам.

Материалы и методы

Существующие методы генерации ответов на естественном языке характеризуются усредненностью и неточностью. Это делает использование этих методов непригодным для передачи обучаемым точных научных сведений. В основе предоставления учебных материалов должны быть традиционные формулировки и положения учебной дисциплины. Перспективным применением диалога на естественном языке представляется применение ранжирующих чат-ботов, выполняющих анализ вопроса обучаемого для предоставления наиболее подходящего и точного ответа, выделенного из содержания дистанционного курса.

Для организации диалога можно использовать стандартные средства создания чат-бота, который будет вести диалог с обучаемыми по содержанию дистанционного курса. Ключевым в организации такого диалога является формирование стандартных ответов на вопросы обучаемых. Роль искусственного интеллекта заключается в сопоставлении вопроса пользователя с эталонами вопросов, выборе наиболее близкого эталона и выдаче ответа, соответствующего эталону.

Для выдачи ответа на вопрос обучаемого в [7] предлагается следующий сценарий:

- 1) Ввод пользователем текстового вопроса по теме онлайн-курса на естественном языке.
- 2) Проведение лексической обработки (разбиение на конструкции и слова, а также их обработка – токенизация и лемматизация).
- 3) Перевод текстовых данных в математическую модель (векторизация с использованием модели word2vec).
- 4) Семантический анализ обработанного вопроса (классификация темы вопроса и поиск релевантных ответов).
- 5) Выдача пользователю ответов в виде топ-3 ответов, наиболее релевантных заданному вопросу.

Используя эту методику, система выдает точный, заранее подготовленный ответ. Методическое обеспечение такой технологии потребует структуризации сведений в виде наборов тем, типовых вопросов и соответствующих готовых ответов. Кроме этого, по этим данным потребуется выполнить лексическую обработку, построение модели и настройку классификатора. Обучение классификатора является наиболее проблематичным, так как требует наличия значительного количества типовых вопросов для одного ответа.

Предлагается заменить определение класса вопроса сравнением его с эталонными вопросами для заранее сформированного набора ответов. Формирование набора ответов может быть выполнено по содержанию курса. Традиционно дистанционный курс делится на темы и лекции. Лекции делятся на страницы достаточно компактного содержания. Страница может служить образцом ответа, а заголовки тем, лекций и страниц – эталонными вопросами. Еще одним источником ответов могут служить словари дистанционных курсов. Термин соответствует эталонному вопросу, а его определение – ответу. Часть тестовых заданий также можно использовать в качестве ответов. Текст задания будет эталонным вопросом, а ответ будет составлен из правильных вариантов.

Предложенная методика не всегда будет давать приемлемые ответы. Заголовки могут не соответствовать возможным вопросам к содержанию страницы, страница может содержать ответы на несколько вопросов. Искусственный интеллект оценивает близость вопроса к эталону с определенной погрешностью. Для верификации распознавания в формируемый ответ целесообразно включить эталонную формулировку вопроса, чтобы пользователь мог сопоставить эталон со своей формулировкой.

Предлагаемая технология нуждается в методической доработке и настройке. До такой настройки она будет выполнять роль путевода по содержанию дистанционного курса. Доработка может быть выполнена по результатам сопоставления вопросов и эталонов. Вопросы, для которых не выбран ответ, либо могут относиться к нераспознанным вопросам, либо к вопросам, на которые нет ответа в курсе. В первом случае можно дополнить систему выбором из списка наиболее релевантных эталонов или выбором по содержанию курса. Сопоставленная пара «вопрос-ответ» будет запоминаться в системе как новый эталон. Во втором случае вопросы без ответа могут использоваться для дополнения курса новыми ответами. Такая методика приведет к адаптации курса к новым вопросам обучаемых.

Диалог может вести чат-бот или пользователь. Активный диалог может начинаться чат-ботом с описания содержания курса, целей, задач и результатов обучения. Предложения чат-бота могут быть основаны на не пройденных пользователем лекциях, не выполненных заданиях или тестах в порядке, установленном содержанием курса. Если пользователя не устраивают предложения чат-бота, то он переводит диалог в режим ответа на свои вопросы – чат-бот будет отвечать на вопросы на основании сравнения вопросов пользователя с эталонами.

Ключевым в предлагаемой технологии является определение сходства вопроса с эталонами. Известно несколько методов решения такой задачи [8, 9]. Технология «мешок слов» в качестве вектора параметров текстов определяет частотные характеристики входящих в текст термов – слов или словосочетаний в некоторой стандартной форме. Предложено несколько таких характеристик:

- индикаторы включений термов в текст;
- абсолютные частоты – количества включений термов в текст;
- относительные частоты – нормализация абсолютных частот на размер текста (tf – term frequency);
- tf-idf – произведение относительной частоты терма и обратной частоты документа (inverse document frequency);
- Bag-of-terms – перечисленные выше характеристики, рассчитанные только по ключевым термам, включенным в словарь.

В результате для корпуса текстов получается сильно разреженная матрица, в которой каждый текст корпуса характеризуется вектором с количеством ненулевых элементов, равным количеству термов в тексте.

В работе [10] для снижения размерности матрицы вместо терма предлагается использовать кластер, объединяющий семантически близкие термы. Семантическую близость предлагается определять на основе векторного описания терма – эмбединга [11]. Векторы термов вычисляются нейронной сетью по корпусу документов и отражают одинаковость контекста их использования. Предложенная модель получила название word2vec. Ее используют для определения близости термов, предсказания терма по контексту или контекста для терма. Для сравнения предложений – цепочек термов – предлагается модель doc2vec, которая определяет эмбединг предложения через усреднение эмбедингов термов, входящих в предложение.

Сравнение вопроса пользователя с эталоном по технологии «мешок слов» возможно на основе косинусной меры близости векторов предложений. Вектор параметров вопроса включает индикаторы вхождения терма в вопрос. Выбор такой характеристики оправдан тем, что вопрос является коротким предложением, в котором значимости слов априори равны. Если вопрос и эталон содержат разные термы, то мера близости будет равна нулю, даже если в вопросе используются синонимы. Такое сравнение отличается излишней «жесткостью».

Составление и использование списков синонимов является достаточно трудоемким. В работе [12] предлагается включать в «мешок слов» запроса «синонимы» – термы из словаря word2vec, близкие к термам запроса по косинусной мере. Такой метод расширяет список документов, релевантных исходному запросу. Для поставленной задачи поиска эталона по вопросу пользователя желательно сузить поиск. Поэтому предлагается использовать меры близости термов вопроса и эталона, определенные на основе модели word2vec или другой модели, которая вычисляет меру близости термов.

Для определения близости вопроса эталону предлагается следующий алгоритм. Строятся два вектора одинаковой размерности: один для эталона, второй для вопроса. Определяются попарные меры близости термов эталона и вопроса. Выбирается пара с максимальной мерой. Для эталона компонента вектора, соответствующая терму, принимается равной единице, а аналогичная компонента вектора вопроса принимается равной мере близости. Далее эта процедура повторяется для несопоставленных термов эталона и вопроса. Если термов в эталоне меньше, то несопоставленные компоненты вектора вопроса приравниваются к единицам, а вектор эталона дополняется нулями. Если термов в эталоне больше, то несопоставленные компоненты вектора вопроса дополняются нулями. Мера близости рассчитывается как косинус угла двух полученных векторов.

В Таблице 1 приведено сравнение вопроса пользователя «структуры данных в компьютерной обработке» с эталоном «технология баз данных в информационных системах». Фоном в таблице выделены максимальные меры близости термов. Разное количество термов в вопросе и эталоне привело к отсутствию сопоставления для терма «информационный». В целом косинусная мера близости векторов эталона и вопроса оказалась достаточно большой (0,860), несмотря на совпадение только одного терма.

Таблица 1 – Сравнение термов эталона и термов вопроса
Table 1 – Comparison of standard terms and question terms

Термы эталона	Термы вопроса				Вектор эталона	Вектор вопроса	Термы вопроса, сопоставленные термам эталона
	структура	данные	компьютерный	обработка			
технология	0,364	0,433	0,720	0,489	1	0,720	компьютерный
база	0,485	0,534	0,504	0,566	1	0,566	обработка
данные	0,463	1,000	0,507	0,587	1	1,000	данные
информационный	0,454	0,300	0,620	0,336	1	0	
система	0,479	0,360	0,531	0,380	1	0,479	структура

Предложенную технологию сопоставления текста с эталоном можно применять для проверки знания определений и положений. Аналогом такой проверки являются тестовые задания в открытой форме. В компьютерных системах тестирования такие задания предполагают ответ в виде числа, слова или буквы. Применение более сложных вариантов таких заданий, например, на знание определений, затруднено по причинам возможных перестановок слов, использования синонимов и применения нескольких вариантов определения сложных объектов. Например, известно много определений таких понятий, как информационная система. Компьютерные системы тестирования обладают излишней жесткостью – небольшое отклонение от образца приводит к неудовлетворительной оценке.

Некоторые авторы предлагают методики «мягкого тестирования» [13]. Такие методики требуют трудоемкой подготовки в виде списка и ранжирования вариантов. В качестве альтернативы предлагается использовать меры близости ответа на тест с

эталонными правильными ответами. Мера близости могла бы заменить традиционную дихотомическую оценку 0 или 1 значением близости от 0 до 1.

Такая технология приблизила бы проверку усвоенного материала к контрольному опросу преподавателя. Диалог проверки знаний можно использовать не только для выставления оценки, но и для определения пробелов и рекомендации материалов для их устранения. Такое применение превращает диалог с чат-ботом в тренажер: обучаемый в диалоге изучает материалы, проходит контрольный опрос, по результатам которого либо дообучается при выявлении ошибок, либо переходит к следующей теме при успешных ответах на вопросы. При этом реализуется основной принцип асинхронного обучения – обучение происходит с учетом индивидуальных возможностей обучаемого.

Предложенная технология сравнения вопроса с эталонами может использоваться для решения разных задач, например, для информирования об организации учебного процесса. При наличии типизированных структур хранения таких данных в реляционных базах данных или JSON-структурах можно выделить эталонные вопросы и соответствующие процедуры извлечения данных. Работа [14] содержит один из первых вариантов решения этой задачи. В частности, по результатам обработки вопроса на естественном языке пользователю демонстрируется формальная постановка. В предлагаемой технологии определения эталона по вопросу пользователя можно демонстрировать пользователю список эталонов, отсортированных по убыванию меры сходства. Выбранный пользователем вариант можно запоминать как дополнительный вариант эталонного вопроса.

Результаты

Ключевым в предлагаемой технологии применения чат-ботов для дистанционного обучения является распознавание семантики вопроса пользователя. Для проверки близости двух текстов использовались модель `doc2vec` из библиотеки `gensim`¹ и предложенный алгоритм, использующий косинусную меру близости эмбедингов термов. Эмбединги определялись с помощью модели `word2vec` той же библиотеки. Для учета специфики курса не использовались модели, полученные по одному из корпусов текстов на русском языке. Вместо этого модель обучалась по пособиям в области информационных технологий.

Результат сравнения некоторых вопросов с эталоном «технология баз данных в информационных системах» приведен в Таблице 2. Предложенный алгоритм дает предсказуемые результаты даже в случае использования разных терминов в близких по смыслу предложениях. Модель `doc2vec` формирует завышенные меры близости с некоторым случайным разбросом. Это объясняется тем, что алгоритм `doc2vec` включает случайный выбор на начальном этапе определения эмбединга предложения. В результате мера близости двух одинаковых предложений (первая строка таблицы) близка к единице, но не равна ей.

Таблица 2 – Меры близости вопросов к эталону
Table 2 – Measures of questions' similarity to the standard

Вопрос для сравнения с эталоном «технология баз данных в информационных системах»	Мера близости по предложенному алгоритму	Мера близости <code>doc2vec</code>
технология баз данных в информационных системах	1	0,952
технология баз данных	0,707	0,841

¹ Gensim Tutorial. URL: <https://www.tutorialspoint.com/gensim/index.htm> (дата обращения: 24.05.2024)

структуры данных в компьютерной обработке	0,860	0,987
применение структур информации	0,705	0,780
группировка сортировка агрегирование показателей	0,522	0,856

В Таблице 3 приведен пример сравнения ответов на тестовое задание в открытой форме. Рассмотрены разные варианты продолжения определения «Отношение находится в 3нф, если...» в сравнении с эталонным «отношение находится в 2нф и каждый неключевой атрибут нетранзитивно зависит от ключа». В таблице ответы расположены по убыванию точности определения и мера близости отражает такое убывание.

Таблица 3 – Меры близости ответов к эталону
Table 3 – Measures of response similarity to the standard

Варианты ответов	Мера близости по предложенному алгоритму	Мера близости doc2vec
отношение находится в 2нф и каждый неключевой атрибут нетранзитивно зависит от ключа	1	0,981
каждый неключевой атрибут нетранзитивно зависит от ключа	0,745	0,951
каждый атрибут нетранзитивно зависит от ключа	0,667	0,964
нет транзитивных зависимостей от ключа	0,644	0,972
отношение сконструировано правильно	0,467	0,885

По результатам измерения мер близости можно сделать следующие выводы. Меры близости, вычисленные с помощью модели doc2vec, не дают в данном случае надежной оценки и в большинстве случаев оказываются завышенными. Предложенный алгоритм сравнения определяет ненулевую косинусную меру, даже для неправильных ответов.

Обсуждение

Предложенный алгоритм сравнения дает более точные результаты по сравнению с doc2vec, но также нуждается в настройке, так как даже далекие от образца варианты определения оценивает, как достаточно близкие. Улучшить предлагаемую методику можно введением некоторых пороговых значений как в отношении мер близости термов, так и в отношении мер близости предложений.

Как и любые технологии использования естественного языка, предложенные методики применения чат-бота для ранжирования ответов на вопрос пользователя и оценки соответствия текстов эталонным образцам характеризуются определенным уровнем неточности, нуждаются в настройке и адаптации. В соответствующей информационной системе должны храниться списки соответствий вопросов и ответов. Такие списки могут, во-первых, использоваться для оценки точности, а во-вторых, пополняться по результатам выбора пользователей, обеспечивая таким образом постоянное дообучение системы.

Запоминание просмотренных пользователем материалов и результатов опросов можно использовать для оценки прогресса обучения и рекомендации еще неизученных материалов.

Заключение

Предложенная методика определения соответствия коротких текстов учитывает не только полное совпадение термов предложений, но и их семантическую близость. В результате достигается большая точность оценки соответствия. Дальнейшее ее повышение возможно за счет обучения модели по расширенному корпусу текстов в комбинации с использованием словарных термов.

Применение метрик соответствия коротких текстов позволяет организовать диалог с обучаемыми на естественном языке для выбора учебных материалов дистанционного курса и для контроля знания базовых определений и положений.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Драндров Д.А., Драндров Г.Л. Плюсы и минусы дистанционного обучения. *Современные проблемы науки и образования*. 2022;(3). <https://doi.org/10.17513/spno.31756>
Drandrov D.A., Drandrov G.L. Pros and cons of distance learning. *Sovremennye problemy nauki i obrazovaniya = Modern problems of science and education*. 2022;(3). (In Russ.). <https://doi.org/10.17513/spno.31756>
2. Киуру К.В., Попова Е.Е., Маковецкая Ю.Г. Новые технологии дистанционного обучения в системе высшего и дополнительного профессионального образования. *Проблемы современного педагогического образования*. 2022;(75-3):196–199.
Kiuru K.V., Popova E.E., Makovetskaya Ju.G. New technologies of distance learning in the system of higher and additional professional education. *Problemy sovremennogo pedagogicheskogo obrazovaniya = Problems of modern pedagogical education*. 2022;(75-3):196–199. (In Russ.).
3. Ibna Riza A.N., Hidayah I., Santosa P.I. Use of Chatbots in E-Learning Context: A Systematic Review. In: *2023 IEEE World AI IoT Congress (AIoT), 07-10 June 2023, Seattle, WA, USA*. IEEE; 2023. P. 0819–0824. <https://doi.org/10.1109/AIoT58121.2023.10174319>
4. Kumar J.A. Educational chatbots for project-based learning: investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education*. 2021;18(1). <https://doi.org/10.1186/s41239-021-00302-w>
5. Labadze L., Grigolia M., Machaidze L. Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*. 2023;20(1). <https://doi.org/10.1186/s41239-023-00426-1>
6. Горячкин Б.С., Галичий Д.А., Цапий В.С., Бурашников В.В., Крутов Т.Ю. Эффективность использования чат-ботов в образовательном процессе. *E-Scio*. 2021;(4):529–551.
Goryachkin B.S., Galichii D.A., Tsapii V.S., Burashnikov V.V., Krutov T.Yu. Effektivnost' ispol'zovaniya chat-botov v obrazovatel'nom protsesse. *E-Scio*. 2021;(4):529–551. (In Russ.).
7. Рожкин П.А., Нехаев И.Н., Маркин К.А. Конструирование системы интеллектуального поиска ответов на вопросы обучающихся на онлайн-курсе на основе word2vec. *International Journal of Advanced Studies*. 2018;8(1):106–128. <https://doi.org/10.12731/2227-930X-2018-1-106-128>
Rozhkin P.A., Nekhaev I.N., Markin K.A. Designing AI teacher assistant on online-course based on word2vec technology. *International Journal of Advanced Studies*. 2018;8(1):106–128. (In Russ.). <https://doi.org/10.12731/2227-930X-2018-1-106-128>

8. Бенгфорт Б., Билбро Р., Охеда Т. *Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка*. Санкт-Петербург: Питер; 2019. 368 с.
Bengfort B., Bilbro R., Ojeda T. *Applied Text Analysis with Python. Enabling Language-Aware Data Products with Machine Learning*. Saint Petersburg: Piter; 2019. 368 p. (In Russ.).
9. Пархоменко П.А., Григорьев А.А., Астраханцев Н.А. Обзор и экспериментальное сравнение методов кластеризации текстов. *Труды Института системного программирования РАН*. 2017;29(2):161–200. [https://doi.org/10.15514/ISPRAS-2017-29\(2\)-6](https://doi.org/10.15514/ISPRAS-2017-29(2)-6)
Parhomenko P.A., Grigorev A.A., Astrakhansev N.A. A survey and an experimental comparison of methods for text clustering: application to scientific articles. *Trudy Instituta sistemnogo programmirovaniya RAN = Proceedings of the Institute for System Programming of the RAS*. 2017;29(2):161–200. (In Russ.). [https://doi.org/10.15514/ISPRAS-2017-29\(2\)-6](https://doi.org/10.15514/ISPRAS-2017-29(2)-6)
10. Юферов В.И., Разин Н.А. Векторизация текстов на основе word-embedding моделей с использованием кластеризации. *Моделирование и анализ информационных систем*. 2021;28(3):292–311. <https://doi.org/10.18255/1818-1015-2021-3-292-311>
Yuferev V.I., Razin N.A. Word-embedding Based Text Vectorization Using Clustering. *Modelirovanie i analiz informatsionnykh sistem = Modeling and Analysis of Information Systems*. 2021;28(3):292–311. (In Russ.). <https://doi.org/10.18255/1818-1015-2021-3-292-311>
11. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. URL: <https://arxiv.org/abs/1301.3781> [Accessed 24th March 2024].
12. Атаева О.М., Серебряков В.А., Тучкова Н.П. О модели поиска синонимов. *Электронные библиотеки*. 2021;24(6):1006–1022.
Ataeva O.M., Serebriakov V.A., Tuchkova N.P. On the synonym search model. *Elektronnye biblioteki = Russian Digital Libraries Journal*. 2021;24(6):1006–1022. (In Russ.).
13. Морев И.А. *Образовательные информационные технологии. Часть 2. Педагогические измерения*. Владивосток: Издательство Дальневосточного университета; 2004. 174 с.
Morev I.A. *Obrazovatel'nye informatsionnye tekhnologii. Chast' 2. Pedagogicheskie izmereniya*. Vladivostok: Izdatel'stvo Dal'nevostochnogo universiteta; 2004. 174 p. (In Russ.).
14. Codd E.F. et al. RENDEZVOUS Version 1: An Experimental English Language Query Formulation System for Casual Users of Relational Data Bases. *IBM Research Report*. 1978;RJ2144.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Братищенко Владимир Владимирович, кандидат физико-математических наук, доцент кафедры математических методов и цифровых технологий Байкальского государственного университета, Иркутск, Российская Федерация.

e-mail: vbrat56@mail.ru

ORCID: [0000-0002-7755-4170](https://orcid.org/0000-0002-7755-4170)

Vladimir V. Bratishchenko, PhD in Physics and Mathematical Sciences, Associate Professor of the Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, the Russian Federation.

*Статья поступила в редакцию 29.05.2024; одобрена после рецензирования 30.05.2024;
принята к публикации 05.06.2024.*

*The article was submitted 29.05.2024; approved after reviewing 30.05.2024;
accepted for publication 05.06.2024.*