

Научная статья

УДК 378 + 004

DOI: 10.24412/2072-9014-2025-373-41-49

КОНЦЕПЦИЯ ИЗУЧЕНИЯ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА (NLP) В ВЫСШЕЙ ШКОЛЕ

Пикалов Иван Юрьевич

Курский государственный университет,
Курск, Россия

ivan.pikalov@gmail.com, <https://orcid.org/0000-0002-5438-7429>

Аннотация. В статье рассматривается концепция изучения методов обработки естественного языка (*англ.* Natural Language Processing, NLP) в высшей школе, позволяющая дать системное представление о данной технологии, показать возможные области использования и инструменты для ее применения. Предлагаемый подход применим для обучения студентов самых различных направлений подготовки. Он позволяет сформировать компетенции в области использования методов искусственного интеллекта для обработки данных на естественном языке для решения профессиональных задач.

Ключевые слова: анализ данных; большие данные; обработка естественного языка; обработка текста; искусственный интеллект; NLP.

Для цитирования: Пикалов И. Ю. Концепция изучения обработки естественного языка (NLP) в высшей школе / И. Ю. Пикалов // Вестник МГПУ. Серия «Информатика и информатизация образования». 2025. № 3 (73). С. 41–49. <https://doi.org/10.24412/2072-9014-2025-373-41-49>

Original article

UDC 378 + 004

DOI: 10.24412/2072-9014-2025-373-41-49

THE CONCEPT OF LEARNING PROCESSING NATURAL LANGUAGE PROCESSING (NLP) IN HIGH SCHOOL

Ivan Yu. Pikalov

Kursk State University,
Kursk, Russia

ivan.pikalov@gmail.com, <https://orcid.org/0000-0002-5438-7429>

Abstract. The article presents the concept of studying Natural Language Processing methods in higher education which allows to give a systematic understanding of this technology, to show possible areas of use and tools for its application. The proposed approach is applicable for teaching students in a wide variety of fields of study. It allows you to form competencies in using Artificial Intelligence methods for data processing in natural language to solve professional tasks.

Keywords: data analysis; big data; natural language processing; text processing; artificial intelligence; NLP.

For citation: Pikalov I. Yu. The concept of learning processsing Natural Language Processing (NLP) in higher education / I. Yu. Pikalov // MCU Journal of Informatics and Informatization of Education. 2025. № 3 (73). P. 41–49. <https://doi.org/10.24412/2072-9014-2025-373-41-49>

Введение

Обработка естественного языка (NLP) является одной из технологий искусственного интеллекта (ИИ). Она используется для работы с большим объемом текстовых данных и находит все более широкое практическое применение. Тот факт, что для ее использования требуется соответствующая подготовка специалистов, позволяет говорить о необходимости ее изучения в высшей школе.

В последнее время стали появляться работы, связанные с изучением технологий ИИ, применяемых для работы с текстовыми данными. В статье В. И. Глизбурга [1] рассматривается применение средств генеративного ИИ в процессе подготовки бакалавров и магистров педагогического образования. Обозначены формируемые компетенции, указана необходимость учета возможных рисков и корректировка методических систем обучения, позволяющих сочетать естественный и искусственный интеллект.

В работе О. Ф. Брыксиной [2] указывается, что современный этап цифровизации образования в большей степени связан с активным внедрением сквозных

технологий и инновационных цифровых решений, а также приводится анализ практических примеров их внедрения в учебные дисциплины. При этом акцент делается как на содержательных, так и на инструментальных аспектах их применения. Технологии ИИ должны выступать и как объект изучения, и как инструменты осуществления профессиональной деятельности.

В работах ряда исследователей рассматривается применение технологий ИИ при обучении иностранным языкам в вузах [3–5]. Основное внимание при этом уделяется следующим вопросам:

- технологии ИИ, используемые (или которые можно использовать) в образовательном процессе;
- ресурсы, содержащие рассматриваемые технологии, описание их возможностей и способов применения;
- описание преимуществ использования рассматриваемых технологий в образовательном процессе и профессиональной деятельности, примеры сценариев их применения;
- возможные риски и возникающие сложности при внедрении технологий ИИ.

При этом под образовательным процессом понимается не только обучение студентов, но и подготовка методических материалов, оценивание отдельных работ, анализ усвоения учебного курса, определение уровня знаний студента, формирование индивидуализированного учебного плана, определение содержания отдельного курса и образовательной программы в целом.

Наиболее востребованными по данному направлению технологиями являются: технология распознавания речи, машинный перевод, технологии обработки естественного языка (NLP), технологии чат-ботов, технология распознавания лиц.

В исследовании Р. Р. Зариповой рассматривается использование цифровых инструментов обработки естественного языка для количественного и статистического анализа языковых данных, необходимого для составления методических рекомендаций по содержанию и результатам освоения программ по иностранному языку для обучения будущих учителей-предметников [6].

В исследованиях подчеркивается, что применение технологий ИИ способствует повышению эффективности процесса обучения. Предлагаются конкретные инструменты, использующие подобные технологии, такие как чат-боты, обучающие системы и специализированные тематические платформы. В работах поднимается вопрос о необходимости формирования специальных умений и навыков владения цифровыми образовательными инструментами.

Однако почти все работы, описывающие возможности отдельных технологий в обучении или профессиональной деятельности и содержащие примеры их использования для решения конкретных задач, не дают системного представления о технологии. При современном развитии технологий ИИ и его темпах роста необходимо формировать системные знания по данному направлению.

Приведем несколько доказательств этого утверждения:

1. Внимание со стороны государства: в 2019 году был принят Указ Президента РФ № 490, направленный на ускоренное внедрение ИИ в различные сферы деятельности¹. В указе говорится не только об увеличении доли работников, имеющих навыки использования технологий ИИ, но и о росте доверия граждан к технологиям ИИ. Это подразумевает не только получение результатов, но и понимание, на чем данные технологии основаны, осознанное применение тех или иных инструментов и алгоритмов и по возможности их настраивание под конкретную задачу.

2. Цифровая компетентность педагога выступает сегодня как стратегическая государственная задача, которая определяется запросами общества и зафиксирована нормативными документами высшего образования.

3. Зная принципы функционирования и область применения конкретной технологии, можно более эффективно использовать ее при решении различных профессиональных задач.

4. Понимая принципы работы технологии, пользователь будет больше доверять полученным результатам, а не относиться к ним как к результатам работы черного ящика.

5. Подготовка в высшей школе нацелена в том числе на формирование системного мышления. В ФГОС высшего образования уровня магистратуры в обязательном порядке присутствует такая категория универсальных компетенций, как системное и критическое мышление.

Методы исследования

В рамках исследования был проведен анализ имеющихся научных результатов в сфере методов обработки естественного языка и их использования для повышения эффективности интеллектуальных систем во всех видах профессиональной деятельности, а также анализ их изучения в высшей школе. Кроме того, использовались такие теоретические методы исследования, как обобщение и систематизация знаний.

Для формирования системных знаний в данной области необходимо понимать структуру изучаемой технологии, ее базовые алгоритмы и принципы, нужно выработать общее представление о данном направлении ИИ, показать уже сформировавшиеся области использования и обозначить возможные направления развития. Кроме того, нужно вырабатывать навыки использования инструментов, пробовать создавать свои проекты по анализу данных, знать успешные примеры решения конкретных профессиональных задач.

В ходе исследования были также проанализированы существующие подходы к изучению технологий обработки естественного языка специалистами

¹ URL: <https://base.garant.ru/72838946> (дата обращения: 14.11.2024).

разных направлений подготовки в высшей школе, обзор современных систем анализа больших данных. Анализ показал, что в настоящее время не все аспекты, необходимые для формирования системного и целостного представления о технологии, рассматриваются в процессе обучения. В основном внимание уделяется применению ресурсов, содержащих рассматриваемые технологии, описанию их возможностей, вариантов использования, преимуществ и недостатков.

Результаты исследования

Результатом работы стала выработка общей концепции изучения обработки естественного языка, которая не только показывает использование конкретных методов и инструментов, но и позволяет рассмотреть данную технологию системно: понять ее структуру, сформировать общее представление о данном направлении ИИ, обозначить область ее применения, показать примеры использования при решении конкретных профессиональных задач.

Для достижения поставленных целей мы предлагаем при изучении обработки естественного языка использовать тематические блоки, показанные на рисунке.



Источник: составлено автором.

Рис. Тематические блоки для изучения обработки естественного языка

Рассмотрим кратко каждый из обозначенных блоков.

Решаемые задачи. Нужно сделать обзор типичных задач, основанных на технологии обработки естественного языка. К таким задачам относятся: машинный перевод, отражение содержания (суммаризация), анализ тональности, распознавание и синтез речи, определение именованных сущностей, классификация текстов, поиск обоснования, извлечение фактов и отношений.

Используемые алгоритмы. Даже если рассматривать только типичные задачи, связанные с обработкой естественного языка, то можно заметить, что они являются достаточно сложными и для их решения используются разные методы. Таких методов немало, и они существенно различаются по сложности, используемым технологиям и реализации. Кроме того, происходит постоянное улучшение существующих алгоритмов и появление новых. Но для понимания принципов работы применяемых технологий достаточно остановиться на простых базовых алгоритмах, которые лежат в основе решения многих востребованных задач. Понимание алгоритмов позволит с большим доверием относиться к результатам, выбирать наиболее подходящий метод для решения конкретной задачи, дополнять или изменять используемые словари или другие ресурсы.

При рассмотрении выбранных алгоритмов необходимо указывать также области применения, подчеркивать их особенности и ограничения. Для формирования общего представления об основных алгоритмах, лежащих в основе многих задач, можно рассмотреть следующие алгоритмы и методы:

- токенизация (любая задача анализа текстов);
- лемматизация (любая задача анализа текстов);
- стемминг (исправление орфографических ошибок в токенах, задачи, в которых скорость и производительность имеют большое значение);
- удаление шумовых слов (стоп-слов);
- векторное представление текста (машинное обучение, большие языковые модели, семантический поиск); желательно рассмотреть несколько конкретных концепций получения векторов, например Word2Vec;
- мешок слов (классификация, определение схожести текстов, машинное обучение);
- глубокое обучение (машинный перевод, синтез и распознавание речи).

Области применения. Данный блок помогает разобраться с возможностями изучаемой технологии, а также показывает, что она является сквозной и рассчитана не только на решение отдельных задач, а должна использоваться как эффективный инструмент, применяемый в самых разных областях. Предлагается обозначить обширные сферы применения с указанием решаемых задач и применяемых технологий. Основное внимание следует уделить описанию тех сфер использования, где нужны определенные компетенции, позволяющие применять рассматриваемые технологии:

– *Маркетинг.* Обработка естественного языка используется для анализа отзывов клиентов, чтобы увидеть сильные и слабые стороны продукта или услуги и понять, как их улучшить.

– *Построение диалоговых систем.* К таким системам относятся умные помощники, чат-боты, QA-системы. Все они опираются на NLP-инструменты: выделение смысла, контекста, определение намерения, а затем выстраивание диалога. При этом некоторые системы могут использовать распознавание и синтез речи. На основе NLP работают многочисленные инструменты

для генерации текстов. Они отвечают на вопросы пользователей, генерируют тексты на разнообразные темы, учитывая разные стили и другие ограничения.

– *Инвестиции*. NLP может извлекать информацию, факты, сущности и другие паттерны из текстов, которые дают понять, что происходит с компаниями, организациями, людьми, определяет возможных стейкхолдеров.

– *Распределение поступающих сообщений*. Можно выделить бинарную классификацию (используется для определения спама или релевантных обращений по определенному направлению) и мультиклассовую. Мультиклассовая классификация может применяться для разделения поступающих обращений в службу поддержки по разным направлениям или для определения тем сообщений.

Конвейер NLP. Реализация любой задачи, как правило, означает построение конвейера обработки (пайплайна), который определяет основные этапы обработки данных. Эти этапы могут отличаться при решении конкретных задач. При этом можно условно выделить две части: предобработку входных данных (обычно занимает больше всего времени) и построение модели. Обозначим семь основных этапов предварительной обработки текстов:

- 1) сегментация и токенизация текста;
- 2) вычисление признаков токенов;
- 3) лемматизация;
- 4) определение значимости и фильтрация стоп-слов;
- 5) разрешение кореференции: в русском и английском языках много местоимений (he, she, it, или ты, я, он и т. д.), которые нужно связать с именами и названиями.

- 6) парсинг зависимостей;

- 7) перевод обработанного текста в векторную форму.

Далее будут следовать этапы, зависящие от конкретной решаемой задачи. Например, построение модели для классификации текстов.

Использование инструментов. На данном этапе изучаются инструменты, использующие рассматриваемые технологии для создания проектов или решения отдельных задач. Здесь возможно использовать готовые инструменты для анализа данных, основанных на этих технологиях, создавать проекты с применением существующих специализированных библиотек и языков программирования и использовать системы анализа данных, работающих по принципу low code и не требующих навыков программирования. Данный этап формирует навыки применения технологий и имеет наибольшую практическую значимость.

Пример реализации подхода, основанного на применении существующих специализированных библиотек и языков программирования показан в работе [7], где предложен подход к изучению методов NLP посредством создания проектов с использованием языка программирования Python и библиотеки SpaCy.

Индивидуальный проект. Для формирования навыков применения технологий наиболее подходящим решением будет выполнение индивидуального проекта по анализу текстовых данных. Важно, что обучающийся, используя полученные знания и умения, будет сам выбирать необходимые технологии и инструменты реализации поставленной задачи.

Заключение

В работе обосновывается необходимость применения системного подхода при изучении технологий обработки естественного языка. Предложенная концепция имеет следующие преимущества:

- 1) может быть использована при обучении студентов разных направлений подготовки, так как не требует углубленных знаний в области ИИ;
- 2) будет способствовать пониманию принципов обработки текстов на естественном языке, что позволит больше доверять полученным результатам;
- 3) обучающиеся получают навыки применения конкретных инструментов, которые можно использовать при решении задач обработки естественного языка в профессиональной деятельности;
- 4) учащиеся рассмотрят различные способы применения изучаемой технологии при решении профессиональных задач.

Список источников

1. Глизбург В. И. Применение искусственного интеллекта при подготовке бакалавров и магистров педагогического образования / В. И. Глизбург // Вестник МГПУ. Серия: Информатика и информатизация образования. 2024. № 2 (68). С. 7–19.
2. Брыксина О. Ф. Сквозные технологии в образовательном процессе: содержательные и инструментальные аспекты / О. Ф. Брыксина // Дистанционные образовательные технологии: материалы VII Междунар. науч.-практ. конф. Симферополь: Ариал. 2022. С. 73–76.
3. Гулая Т. М. Вызовы и перспективы использования технологий искусственного интеллекта в практике преподавания иностранного языка в вузе / Т. М. Гулая, Т. Л. Герасименко // Глобальный научный потенциал. 2024. № 4 (157). Т. 1. С. 38–41.
4. Гарифова О. А. Использование искусственного интеллекта в обучении иностранным языкам студентов неязыковых вузов / О. А. Гарифова, М. А. Хусаинова // Вестник ЮУрГУ. Серия: Образование. Педагогические науки. 2024. Т. 16. № 2. С. 108–116.
5. Верховод А. С. Использование технологии искусственного интеллекта в преподавании иностранного языка / А. С. Верховод // Проблемы современного педагогического образования. 2024. № 83-2. С. 71–73.
6. Зарипова Р. Р. Применение цифровых инструментов обработки естественного языка для оптимизации подготовки учителя-предметника в полилингвальной образовательной среде / Р. Р. Зарипова // Педагогика. Вопросы теории и практики. 2024. Т. 9. Вып. 1. С. 49–55.

7. Пикалов И. Ю. Изучение методов обработки естественного языка с использованием языка программирования Python и библиотеки SpaCy в высшей школе / И. Ю. Пикалов // Вестник МГПУ. Серия: Информатика и информатизация образования. 2024. № 2 (68). С. 48–61.

References

1. Glizburg V. I. The application of artificial intelligence in preparation of bachelors and masters of teacher education / V. I. Glizburg // MCU Journal of Informatics and Informatization of Education. 2024. No. 2 (68). P. 7–19.

2. Bryksina O. F. End-to-end technologies in the educational process: content and instrumental aspects / O. F. Bryksina // Distance educational technologies: materials of the VII International Scientific and Practical Conference. Simferopol: Arial, 2022. P. 73–76.

3. Gulaya T. M. Challenges and prospects of using artificial intelligence technologies in the practice of teaching a foreign language at a higher education institutions / T. M. Gulaya, T. L. Gerasimenko // Global Scientific Potential. 2024. No. 4 (157). V. 1. P. 38–41.

4. Garifova O. A. Use of artificial intelligence in teaching foreign languages to students of non-linguistic universities / O. A. Garifova, M. A. Khusainova // South Ural State University Journal of Education. Educational Sciences. 2024. V. 16. No. 2. P. 108–116.

5. Verkhovod A. S. The use of artificial intelligence technology in teaching a foreign language / A. S. Verkhovod // Problems of modern pedagogical education. 2024. No. 83-2. P. 71–73.

6. Zaripova R. R. Applying digital natural language processing tools to optimize subject teacher training in a multilingual educational environment / R. R. Zaripova // Pedagogy. Theory & Practice. 2024. Vol. 9. No. 1. P. 49–55.

7. Pikalov I. Yu. Learning natural language processing techniques using the Python programming language and the SpaCy library in high school / I. Yu. Pikalov // MCU Journal of Informatics and Informatization of Education. 2024. No. 2 (68). P. 48–61.

Статья поступила в редакцию: 02.06.2025;
одобрена после рецензирования: 04.08.2025;
принята к публикации: 11.08.2025.

The article was submitted: 02.06.2025;
approved after reviewing: 04.08.2025;
accepted for publication: 11.08.2025.

Информация об авторе / Information about author:

Иван Юрьевич Пикалов — кандидат педагогических наук, доцент, доцент кафедры компьютерных технологий и информатизации образования, руководитель научно-методического центра разработки информационных систем и анализа данных, Курский государственный университет, Курск, Россия.

Ivan Yu. Pikalov — Candidate of Pedagogical Sciences, Associate Professor, Associate Professor of the Department of Computer Technology and Informatization of Education, Head of the Center for Information Systems Development and Data Analysis, Kursk State University, Kursk, Russia.

ivan.pikalov@gmail.com, <https://orcid.org/0000-0002-5438-7429>