

ИСПОЛЬЗОВАНИЕ ChatGPT ДЛЯ АВТОМАТИЗАЦИИ ВЫЯВЛЕНИЯ И АНАЛИЗА ТЕКСТОВОЙ ИНФОРМАЦИИ НА C#

А.В. Полуэктов¹, А.И. Заревич¹, Е.А. Попова¹

¹ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

Аннотация. Данная статья рассматривает возможности использования ChatGPT для автоматизации выявления и анализа текстовой информации на языке программирования C#. Будет рассмотрено описание ChatGPT, его применение для обработки текстовой информации и преимущества использования данной технологии.

Ключевые слова: ChatGPT, C#, автоматизация, текстовая информация, анализ, применение, преимущества

USING ChatGPT TO AUTOMATE TEXT INFORMATION DETECTION AND ANALYSIS IN C#

A.V. Poluektov¹, A.I. Zarevich¹, E.A. Popova¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. This article examines the possibilities of using ChatGPT to automate the identification and analysis of text information in the C# programming language. A description of ChatGPT, its application for processing text information and the advantages of using this technology will be considered.

Keywords: ChatGPT, C#, automation, text information, analysis, application, benefits

Обработка и анализ текстовой информации играют важную роль в различных областях, например, бизнес, наука и технологии. Однако, в современном мире объем текстовых данных становится столь огромным, что выполнение операций по обработке и анализу информации без помощи современных технологий становится практически невозможной или очень трудоемким. Рассмотрим возможности использования технологии ChatGPT для автоматизации анализа текстовой информации с использованием языка программирования C#, при этом

ChatGPT позволяет выполнять не только генерацию текста, но и также позволяет осуществить ее обработку по разработанному шаблону действий. При разработке программ на языке программирования C# часто требуется обрабатывать большие массивы текстовых данных при этом в качестве элемента обработки может быть использован ChatGPT который изначально разрабатывался именно для работы с большими массивами текстовой информации.

Разработка технологии ChatGPT началась с модели GPT, аббревиатура с Generative Pre-trained Transformer, разработанная OpenAI в 2018 году. Так как полученная технология оказалась удачной, то появились улучшенные версии данной модели – GPT-2 и GPT-3. Технология ChatGPT основан на модели GPT-3 где в качестве алгоритмов обработки текстовых данных стали следующие направления: чат-боты, системы виртуального ассистента и системы генерации и обработки текстовых данных. Особенность последнего направления строится на системе вопрос-ответ, в котором система на основе огромной базы знаний генерирует текстовую информацию построенную на вопросе пользователя, оформляет ее по шаблону сформированным на принципе подобия с другими документами и отправляет в удобной для получателя форме на языке пользователя. Для взаимодействия языка программирования с ChatGPT была разработана библиотека OpenAI GPT-3 API [2], [3], [7], [8]. Работа ChatGPT состоит из нескольких этапов: построение модели, обработка запроса пользователя, генерация ответа на запрос пользователя.

Опираясь на алгоритмическую конструкцию обработки текстовых данных следует уточнить, следующие параметры – что является параметром запроса, по какому алгоритму или правилу запрос строится, чтобы он мог быть построен программно, как выполняется генерация текста по запросу, чтобы мы могли разработать алгоритмы обработки данных, и как выполняется генерация ответа.

Рассматривая документацию по использованию библиотеки OpenAI GPT-3 API для языка программирования C# можно выделить ответы на поставленные выше вопросы. Так ChatGPT выполняет генерацию текста путём анализа текста запроса и выполняет обработку собственной базы текстов на соответствие слов запроса релевантной информации в своей базе данных. На основании полученных результатов системы выполняет генерацию ответа по выстроенным логическим цепочкам аналогичных запросов. Полученный ответ затем преобразовывается системой в предложенный разработчиком интерфейс пользователя где чаще всего используются или через веб-интерфейс, так и напрямую через программный интерфейс (API). При работе с запросом следует помнить, что при решении

поставленной задачи система ChatGPT разбивает запрос на две составляющие – ключевые слова и логика запроса, ключевые слова отвечают за подбор информации по соответствию текста повторяющимся словам и выражениям, а логика запроса может включать как анализ тональности, так и выявление ключевых проблем, на которые построен запрос. Поэтому, использование технологии ChatGPT в разработке программ позволяет создать систему, способную как обрабатывать большие объемы текста, так и выделять ключевую информацию и обеспечивать аналитику на основе текстовых данных. Это дает программисту более простой подход в построении ПО со сложными алгоритмами обработки текстовой информации.

В качестве среды разработки ПО содержащего большие массивы текстовых данных будем использовать Microsoft Visual Studio и язык программирования C#. Наш выбор на данную среду пал из-за возможности интегрирования различных библиотек, в частности и библиотеку ChatGPT, а также широкими возможностями языка C# позволяющего работать с внешними библиотеками и онлайн системами [1], [5], [6]. Интеграция библиотеки в Microsoft Visual Studio выполняется через команду Install-Package OpenAI.Gpt, где в проект будут импортированы необходимые библиотеки.

Выбор языка C# был нами определен также из-за особенности использования ChatGPT в проектах которая заключается в необходимости ее авторизации. Для этого необходимо сначала провести регистрацию на сайте OpenAI и получить ключ с помощью которого выполняется авторизация. Разработчик получает ключ самостоятельно, при этом пользователю ПО этим заниматься нет необходимости. Авторизация на сайте OpenAI выполняется согласно программному коду, представленному на рисунке 1. Разработчик формирует экземпляр класса Gpt3 и через функцию API передать ключ авторизации на сайт OpenAI. Затем данный экземпляр класса может быть использован для генерации данных. Особенность ChatGPT состоит в том, что в запросе мы указываем условие запроса, как и в любом формальном языке, при этом точность поставленного запроса влияет на получаемый ответ. Для построения запроса используется параметр `prompt`. ChatGPT позволяет также к тексту прилагать такой параметр как "творческий" который дает возможность системе генерации «отходить» от «жестко поставленного вопроса», давая творчество ChatGPT. При этом меняя данный параметр в значениях от 0 до 1 мы можем создать мало предсказуемый текст. В запросе параметр "творческий" задан служебным параметром `temperature`. И последний

параметр, `maxTokens`, определяет количество блоков генерируемого текста. Данный параметр также зависит от условий решаемых задач и подбирается разработчиком ПО индивидуально.

```
var gpt = new Gpt3("your-api-key");  
var response = gpt.Complete(prompt, temperature: 0.7, maxTokens: 150);
```

Рисунок 1 – Использование метода `Complete` для отправки запроса и получения ответа

После выполнения программного кода, рисунок 1, программа сохранит полученные данные в объекте `Response`. При этом объект `Response` представляет собой множество текстовых данных разделенных маркером и общее количество таких записей в объекте определяется параметром `maxTokens`. С каждой записью программист может выполнять различные манипуляции, в частности, вывести их на экран. Для этого можно использовать следующую команду

```
Console.WriteLine(response.Choices[0].Text);
```

На рисунке 2 нами представлен полный код запроса и вывода информации на экран. `C#` с полученной информацией позволяет выполнять не только вывод данных на экран, но и сохранять данные в файл, или передавать другим системам и т.д. Выбор действий над данными зависит только от поставленной задачи [4], [9].

```
var gpt = new Gpt3("your-api-key");  
var prompt = "Once upon a time";  
var response = gpt.Complete(prompt, temperature: 0.7, maxTokens: 150);  
Console.WriteLine(response.Choices[0].Text);
```

Рисунок 2 – Полный код запроса и сохранение ответа с использованием технологии ChatGPT

Выполним оценку актуальности использования средств ChatGPT в системе обработки данных. Рассмотрим на двух программных кодах, представленных на рисунках 3 и 4 систему обработки больших объемов текстовых данных как с использованием собственных средств языка программирования `C#`, так и средствами ChatGPT. Особенность обработки средствами языка программирования `C#` состоит в том, что текст изначально разбивается на элементы динамического

массива. Массив данных для упрощенной обработки затем преобразуется в коллекцию и к ней применяется алгоритм обработки данных, и полученный результат формируется в новый массив данных. Все операции требуют, как больших данных для хранения, так и имеют сложность $O(n)$. При формировании системы анализа текста средствами ChatGPT, программист задает только систему анализа через функцию AnalyzeText и вывод полученного результата. Здесь отпадает необходимость формирования динамических массивов, что уменьшает занимаемую память и упрощает систему обработки данных задач ключевое слово для обработки, а не строя сложную систему запросов [4], [9].

```
string text = "Это текст, который нужно проанализировать на наличие ключевых  
слов";  
string[] words = text.Split(' ');  
Dictionary<string, int> wordCount = new Dictionary<string, int>();  
foreach (string word in words) {  
    if (wordCount.ContainsKey(word)) { wordCount[word]++; }  
    else { wordCount.Add(word, 1); }  
}  
foreach (var pair in wordCount) { Console.WriteLine("{0}: {1}", pair.Key, pair.Value); }
```

Рисунок 3 – Программный код без использования ChatGPT

```
// использование ChatGPT API для анализа текста  
string text = "Это текст, который нужно проанализировать на наличие ключевых  
слов";  
string[] keywords = ChatGPT.AnalyzeText(text);  
foreach (string keyword in keywords) { Console.WriteLine(keyword); }
```

Рисунок 4 – Программный код с использованием ChatGPT

Поэтому, при разработке ПО с большим объемом текстовых данных можно рекомендовать использование как среду разработки Microsoft Visual Studio и язык программирования C#, так и средства ChatGPT для генерации и обработки больших массивов текстовых данных. Кроме выше сказанного ChatGPT позволяет также упростить решение следующих задач:

- проводить анализ больших объемов текстовой информации, выделять ключевые слова текста, задавать тональность и смысловую нагрузку текста;
- использовать интерфейс API для интеграции данных в различные программные продукты и платформы;
- разрабатывать ПО с элементами само обучаемости.

Подводя итог по использованию ChatGPT построения системы автоматизированной обработки текстовой информации с использованием языка программирования C# следует дать более мощный инструмент, который позволяет улучшить процессы обработки и понимания больших объемов текстовых данных. Благодаря чему ChatGPT позволяет эффективно решать множество задач, связанных с обработкой текстов, в том числе автоматическое извлечение информации, анализ тональности, определение тематики и многое другое. Внедрение ChatGPT в процессы автоматизации позволит повысить эффективность работы системы и улучшить качество получаемых результатов.

Список литературы

1. Киселева, А. А. Использование chatGPT в образовании: преимущества и недостатки / А. А. Киселева // Галактика науки-2023, Краснодар, 19–22 апреля 2023 года / Кубанский государственный университет. Т. 4. – Краснодар: Кубанский государственный университет, 2023. – С. 53-57. – EDN TSDTVVF.
2. Тюгинбаев, Д. Д. К вопросу о функционировании языковых моделей: как работает chatgpt / Д. Д. Тюгинбаев // Искусственный интеллект и большие данные (Big data) в судебной и правоохранительной системе: реалии и требование времени: МАТЕРИАЛЫ МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ, Астана, 19 мая 2023 года. – Косшы: Академия правоохранительных органов при Генеральной прокуратуре Республики Казахстан, 2023. – С. 96-99. – EDN UJUKKL.
3. Нургалиев, Т. С. Как работает chatgpt / Т. С. Нургалиев // Первые шаги в науку третьего тысячелетия: Материалы XIX Всероссийской студенческой научно-практической конференции, Нефтекамск, 18 апреля 2023 года. – Уфа: Уфимский университет науки и технологий, 2023. – С. 146-148. – EDN OUMWOA.
4. Степанов, В. К. Возможности применения языковой модели искусственного интеллекта ChatGPT-3.5 в библиотечно-библиографической деятельности / В. К. Степанов, М. Ш. Маджумдер, Д. Д. Бегунова // Научно-техническая информация. Серия 1: Организация и методика информационной работы. – 2023. – № 7. – С. 11-21. – DOI 10.36535/0548-0019-2023-07-2. – EDN UYTXLS.
5. Liferenko, K. V. Pascal and C# programming languages: a comparative analysis / K. V. Liferenko, A. I. Gerasimov // Молодежь. Общество. Современная наука, техника и инновации. – 2022. – No. 21. – P. 55-57. – EDN VTSQFX.

6. Никитин, А. Е. Разработка практикума по изучению объектно-ориентированного языка программирования С# для старших классов / А. Е. Никитин // Актуальные проблемы современной науки: взгляд молодых ученых: Материалы Международной научно-практической конференции, Грозный, 26–27 апреля 2019 года. – Грозный: Чеченский государственный педагогический университет, 2019. – С. 530-534. – EDN ZCVRGH.

7. Белякова, Ю. В. Концепт-параметры как механизм развития средств обобщённого программирования в языке С# / Ю. В. Белякова, С. С. Михалкович // Современные информационные технологии и ИТ-образование. – 2015. – Т. 11, № 2. – С. 205-213. – EDN WAQFBP.

8. Оксюта, О.В. Разработка математической модели оптимального функционирования транспортно-логистического комплекса / О.В.Оксюта, В.А. Коротких // Моделирование систем и процессов. – 2017. – Т. 10, № 3. – С. 55-66.

9. Зольников К.В. Координация проектных работ в области СНК и сложно функциональных блоков / Зольников К.В., Анциферова В.И., Евдокимова С.А., Гречаный С.В. // Моделирование систем и процессов. 2020. Т. 13. № 3. С. 71-76.

10. Полуэктов А.В., Макаренко Ф.В., Ягодкин А.С. Использование сторонних библиотек при написании программ для обработки статистических данных // Моделирование систем и процессов. – 2022. – Т. 15, № 2. – С. 33-41.

References

1. Kiseleva, A. A. Using chatGPT in education: advantages and disadvantages / A. A. Kiseleva // Galaxy of Science-2023, Krasnodar, April 19–22, 2023 / Kuban State University. Volume 4. – Krasnodar: Kuban State University, 2023. – P. 53-57. – EDN TSDTVTF.

2. Tyuginbaev, D. D. On the issue of the functioning of language models: how chatgpt works / D. D. Tyuginbaev // Artificial intelligence and big data (Big data) in the judicial and law enforcement system: realities and the demands of the time: MATERIALS OF THE INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE , Astana, May 19, 2023. – Kossyhy: Academy of Law Enforcement Agencies under the General Prosecutor’s Office of the Republic of Kazakhstan, 2023. – P. 96-99. – EDN UJUKKL.

3. Nurgaliev, T. S. How chatgpt works / T. S. Nurgaliev // First steps into the science of the third millennium: Materials of the XIX All-Russian Student Scientific

and Practical Conference, Neftekamsk, April 18, 2023. – Ufa: Ufa University of Science and Technology, 2023. – P. 146-148. – EDN OUMWOA.

4. Stepanov, V. K. Possibilities of using the language model of artificial intelligence ChatGPT-3.5 in library and bibliographic activities / V. K. Stepanov, M. Sh. Majumder, D. D. Begunova // Scientific and technical information. Series 1: Organization and methodology of information work. – 2023. – No. 7. – P. 11-21. – DOI 10.36535/0548-0019-2023-07-2. – EDN UYTXLS.

5. Liferenko, KV Pascal and C# programming languages: a comparative analysis / KV Liferenko, AI Gerasimov // Youth. Society. Modern science, technology and innovation. – 2022. – No. 21. – P. 55-57. – EDN VTSQFX.

6. Nikitin, A. E. Development of a workshop on studying the object-oriented programming language C# for high school / A. E. Nikitin // Current problems of modern science: the view of young scientists: Proceedings of the International Scientific and Practical Conference, Grozny, April 26–27, 2019 of the year. – Grozny: Chechen State Pedagogical University, 2019. – P. 530-534. – EDN ZCVRGH.

7. Belyakova, Yu. V. Concept parameters as a mechanism for the development of generalized programming tools in the C# language / Yu. V. Belyakova, S. S. Mikhalkovich // Modern information technologies and IT education. – 2015. – T. 11, No. 2. – P. 205-213. – EDN WAQFBP.

8. Oksyuta, O.V. Development of a mathematical model of the optimal functioning of the transport and logistics complex / O.V. Oksyuta, V.A. Korotkikh // Modeling of systems and processes. – 2017. – T. 10, No. 3. – P. 55-66.

9. Zolnikov K.V. Coordination of design work in the field of SNK and complex functional blocks / Zolnikov K.V., Antsiferova V.I., Evdokimova S.A., Grechany S.V. // Modeling of systems and processes. 2020. T. 13. No. 3. P. 71-76.

10. Poluektov A.V., Makarenko F.V., Yagodkin A.S. The use of third-party libraries when writing programs for processing statistical data // Modeling of systems and processes. - 2022. – Vol. 15, No. 2. – pp. 33-41.