

Подводя итоги, можно с уверенностью утверждать, что промпт инженеры нужны для работы с инструментами ИИ. Они выступают в качестве связующего звена между идеей заказчика и реализацией ее генеративной моделью. Вакансии на должность промпт-инженера все чаще встречаются на просторах сети, что говорит о востребованности и перспективах развития профессионалов в данной сфере.

Библиографический список

1. Википедия. 2023. Галлюцинация (искусственный интеллект) // (дата обращения: 26.03.2023) URL: [https://ru.wikipedia.org/wiki/Галлюцинация_\(искусственный_интеллект\)](https://ru.wikipedia.org/wiki/Галлюцинация_(искусственный_интеллект))
2. Сайт нейросети ChatGPT [Электронный ресурс] // URL: <https://openai.com/blog/chatgpt/>
3. Сайт нейросети Dall-e от OpenAI [Электронный ресурс] // URL: <https://openai.com/dall-e-2/>
4. Сайт нейросети Ru-Dall-e от Sber AI [Электронный ресурс] // URL: <https://rudalle.ru/>
5. Co:here. 2022. Prompt Engineering [Электронный ресурс] // URL: <https://docs.cohere.ai/docs/prompt-engineering>.
6. Hunter N. The Art of Prompt Engineering with chatGPT: A Hands-On Guide. 2023 – с.23-25
7. Wikipedia. 2022. Hallucination (artificial intelligence) (дата обращения: 26.03.2023) [Электронный ресурс] / URL: [https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence))

УДК 681.3.06
ГРНТИ 20.19.27

ЧАСТОТНЫЙ СЛОВАРЬ КАК СПОСОБ ИЗУЧЕНИЯ И АНАЛИЗА ТЕКСТОВ

Капличенко В. Ю., Устинов Д. А.

*Трехгорный технологический институт – филиал ТТИ НИЯУ МИФИ,
г. Трехгорный, Челябинская область*

nekomaid421@gmail.com, dima.ustinov2003@yandex.ru

В статье раскрывается использование частотного словаря в различных сферах деятельности, также показано его практическое применение, как инструмента для изучения текста. В данной работе мы подтверждаем гипотезу: при составлении частотного словаря, мы узнаем тематику текста, можем провести его анализ, не читая сам текст.

Ключевые слова: словарь, частотность, список, слова, программа.

FREQUENCY DICTIONARY AS A WAY OF STUDYING AND ANALYZING TEXTS

Kaplichenko V. Y., Ustinov D. A.

TTI NRNU MEPhI, Trekhgorny

The article reveals the use of the frequency dictionary in various fields of activity, and also shows its practical application as a tool for studying the text. In this paper, we confirm the hypothesis: when compiling a frequency dictionary, we learn the subject of the text; we can analyze it without reading the text itself.

Keywords: dictionary, frequency, list, words, program.

Введение. Словарь – источник, содержащий перечень слов, информация в котором упорядочена с помощью разбивки на небольшие статьи, отсортированные по названию или тематике, обычно с пояснениями, толкованиями или переводом на другой язык.

Частотный словарь - вид словаря, в котором приводятся числовые характеристики употребительности слов (словоформ, словосочетаний) какого-либо языка, в том числе языка писателя, какого-либо произведения и т.д. Обычно в качестве характеристики употребительности используется частота встречаемости слова в тексте определенного объема.

Впервые частотные словари появились в отечественной лексикографии в 60-е г. XX века. В 1963 г. в Таллине Э. А. Штейнфельдт издал «Частотный словарь современного русского языка», содержащий 2500 наиболее употребительных слов.

Многие годы авторам учебников по русскому языку для национальных школ «Частотный словарь современного русского языка» Э. А. Штейнфельдта являлся источником определения лексических минимумов [1].

Через некоторое время было проведено новое более серьезное изучение частотной лексики русского языка. За основу были взяты машинные выборки миллиона употреблений слов. В качестве результата данного исследования в 1977 году группой лингвистов под руководством Л.Н. Засориной был издан «Частотный словарь русского языка». Он включает в себя 40 тысяч слов, расположенных по мере убывания употребительности.

В настоящее время применение частотного словаря обладает широким спектром применения. Для работы его могут использовать не только лингвисты, но и представители других различных сфер деятельности.

Сферы применения частотного словаря:

1. *Для изучения иностранного текста, без знания языка.*

В программу можно загрузить текст, написанный на латинском и кириллическом алфавитах, выбрать сортировку по убыванию и запустить. По завершению программа выдаст результат, в котором наиболее часто встречающиеся слова будут находиться на вершине списка. И используя онлайн-переводчик или билингвистический словарь, можно перевести слова, расположенные в начале списка, и по ним понять ключевую мысль текста.

2. *Написание научных статей.*

Исследователи могут воспользоваться частотным словарем при написании научных статей. Загрузив статью в программу и прогнав ее, учащийся может выбрать сортировку по убыванию, в верхнем списке слов он получит ключевые слова его статьи.

3. *Криптоанализ.*

Частотный словарь можно использовать для расшифровки зашифрованных текстов. Этим методом пользовались еще арабы в 14 веке. Ученые заметили, что в зашифрованных текстах некоторые буквы встречаются чаще остальных, сопоставив эти буквы с часто встречающимися буквами из английского алфавита, они смогли расшифровать зашифрованное послание. Этот метод дешифрования используют и по настоящее время.

4. *Анализ текста*

В программу можно загрузить текст, написанный на латинском и кириллическом алфавитах, выбрать сортировку по убыванию и запустить. По завершению программа выдаст результат, в котором наиболее часто встречающиеся слова будут находиться на вершине списка. Таким способом можно проанализировать различные произволения одного автора, выявить наиболее часто встречающиеся слова во всех его произведениях и таким образом определить его стиль написания текстов.

Основная часть. В современном мире большинство предпочитают использовать тексты в электронном формате. Поэтому мы решили написать программу на языке программирования C#, которая будет генерировать частотный словарь. С бумажной версией текста это бы заняло огромное количество времени. Поэтому использование программы поможет минимизировать затрату времени на анализ текста.

Актуальность нашей работы заключается в том, что мы создали программу, которая будет понятна и удобна в использовании любому пользователю, а новизна работы в том, что программное обеспечение (ПО) [5] написано на современном языке программирования, который сейчас набирает обороты в использовании [2], а также в программу можно загружать тексты, на всех языках, использующих в своем написании латинский [3] и кириллический [4] алфавит.

Цель нашей работы заключается в написании ПО для составления частотного словаря, по тексту, загруженному в программу, написанному на латинице или кириллице.

Гипотеза: если мы будем составлять частотный словарь, то мы будем знать тематику теста, не читая его.

При запуске ПО перед пользователем появляется окно, в верхней строке которого нужно указать путь к файлу [6], сделать это можно вручную или же выбрать файл с помощью диалогового окна.

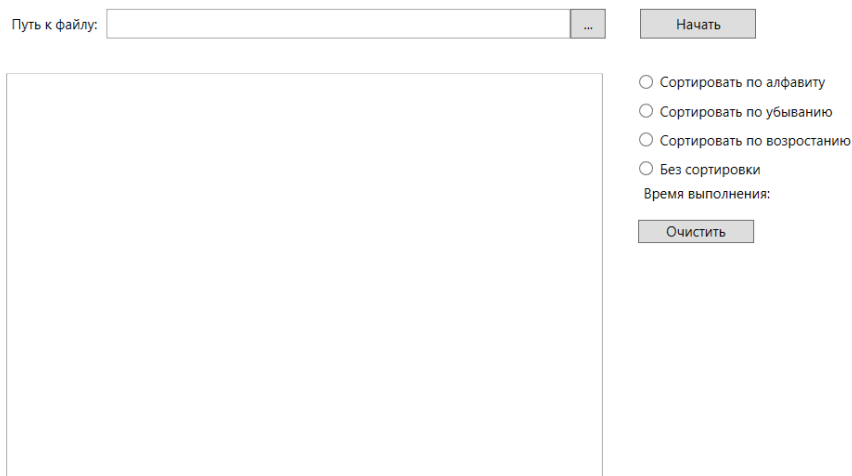


Рисунок 1 – Основное окно программы

После выбора файла по желанию можно указать какой из трёх типов сортировки (по алфавиту, по убыванию, по возрастанию) применить к словарю или же можно вовсе не применять сортировку.

Затем по нажатию кнопки “Начать” программа начнёт обработку текста, и через некоторое время (время выполнения зависит от размера файла) в специальной области появится список слов, отсортированный в соответствии с указанным видом сортировки.

По желанию пользователь может очистить список слов с помощью кнопки “Очистить”

Алгоритм работы программы. По нажатию кнопки “Начать” Вызывается функция `btnStartCounting_Click(object, RoutedEventArgs)`, затем происходит очищение поля списка слов и проверка на корректность пути к файлу. После создаётся список объектов класса `WordQuantity`, который содержит в себе два поля: `string` (для хранения самого слова) и `ushort` (для хранения частоты встречаемости слова в тексте, количество раз). А затем сортируется в зависимости от указанного типа сортировки

Список заполняется с помощью функции `ReadTxt(string)`. Эта функция возвращает неотсортированный список объектов класса `WordQuantity`, в теле этой функции создаётся список всех слов текста “`rawList`”, имеющий тип `string` (этот список также содержит дубликаты слов) и неотсортированный список “`processedList`” типа `WordQuantity`.

Алгоритм перебирает каждое слово (элемент списка “`rawList`”) и проверяет его наличие в “`processedList`”, если слово отсутствует в списке, то оно добавляется в конец. Если же в “`processedList`” уже имеется данное слово, то поле объекта, отвечающее за частоту встречаемости слова в тексте (`Count`) увеличивается на 1.

```
foreach (string wordInList in rawList)
```

```

{
    bool isNewWord = true;
    for (int i = 0; i < processedList.Count; i++)
    {

        if (string.Compare(wordInList, processedList[i].Word) == 0)
        {
            processedList[i].Count++;
            isNewWord = false;
            break;
        }
    }
    if (isNewWord)
    {
        processedList.Add(new WordQuantity(wordInList));
    }
}

```

Создание списка “rawList” происходит путём считывания текста из файла в строку, а последующее её разбиение на список строк с помощью функций `Split(char[], StringSplitOptions)` и `ToList()`

Заключение: Изучая данную тему, мы кратко рассказали об истории появления частотного словаря, об областях его применения и пользе. Также мы написали программу и объяснили принцип ее работы.

В своей работе мы также доказали, что использовать частотный словарь в своей работе могут не только лингвисты, как это принято считать, но и представители других различных сфер деятельности.

Таким образом, мы подтвердили гипотезу: если мы будем составлять частотный словарь, то мы будем знать тематику теста, не читая его.

В заключении хочется еще раз отметить, что использование частотного словаря может сыграть важную роль при анализе текста, ведь его применение может в разы сократить время оценки.

Библиографический список

1. Фридман, Х.Х. О некоторых критериях отбора словаря минимума (СМ) для курса практики устной и письменной речи / Х.Х. Фридман // Вопросы научной организации преподавания иностранных языков в специальном вузе. Пути внедрения результатов научных исследований в практику обучения: тезисы докладов и сообщений научно-теоретической конференции. – Л.: Изд-во ЛГУ, 1970. – С. 144–146.
2. Евдокимов П.В. С# на примерах / Евдокимов П.В. // 1.7 Язык С#. Изд-во «Науки и техника», 2016г – С.24-27.
3. Латиница – [Электронный ресурс] URL: – <https://ru.wikipedia.org/wiki/Латиница> (дата обращения 20.03.2003).
4. Алфавиты на основе кириллицы – [Электронный ресурс] URL: – https://ru.wikipedia.org/wiki/Алфавиты_на_основе_кириллицы (дата обращения 20.03.2023).
5. Программное обеспечение – [Электронный ресурс] URL: – https://ru.wikipedia.org/wiki/Программное_обеспечение
6. Путь к файлу – [Электронный ресурс] URL: – https://ru.wikipedia.org/wiki/Путь_к_файлу