

**Министерство образования Республики Беларусь**

**Учреждение образования  
«Белорусский государственный университет информатики и  
радиоэлектроники»**

**ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И УПРАВЛЕНИЯ**

**Кафедра интеллектуальных информационных технологий**

**Отчет по лабораторной работе №2**

**по курсу «Естественно-языковые интерфейсы интеллектуальных  
систем»**

**на тему «Разработка автоматизированной системы формирования  
словаря естественного языка»**

Выполнили:  
Студенты группы 821702

Астапович О.С.  
Зайцев Н.

Проверил:

Крапивин Ю.Б.

Минск  
2021

**Цель работы:** освоить принципы разработки прикладных сервисных программ для подготовки исходных данных естественно-языковых систем.

Основные задачи работы следующие:

1. Изучить предварительные этапы разработки естественно-языковых систем.
2. Ознакомиться с типологией словарей, используемых в естественно-языковых системах.
3. Освоить на практике процесс разработки автоматизированной системы формирования словаря естественного языка для его последующего использования в рамках естественно-языковой системы.
4. Закрепить навыки программирования естественно-языковых систем и вспомогательных прикладных программ на одном из языков программирования.

### **Вариант задания №5.**

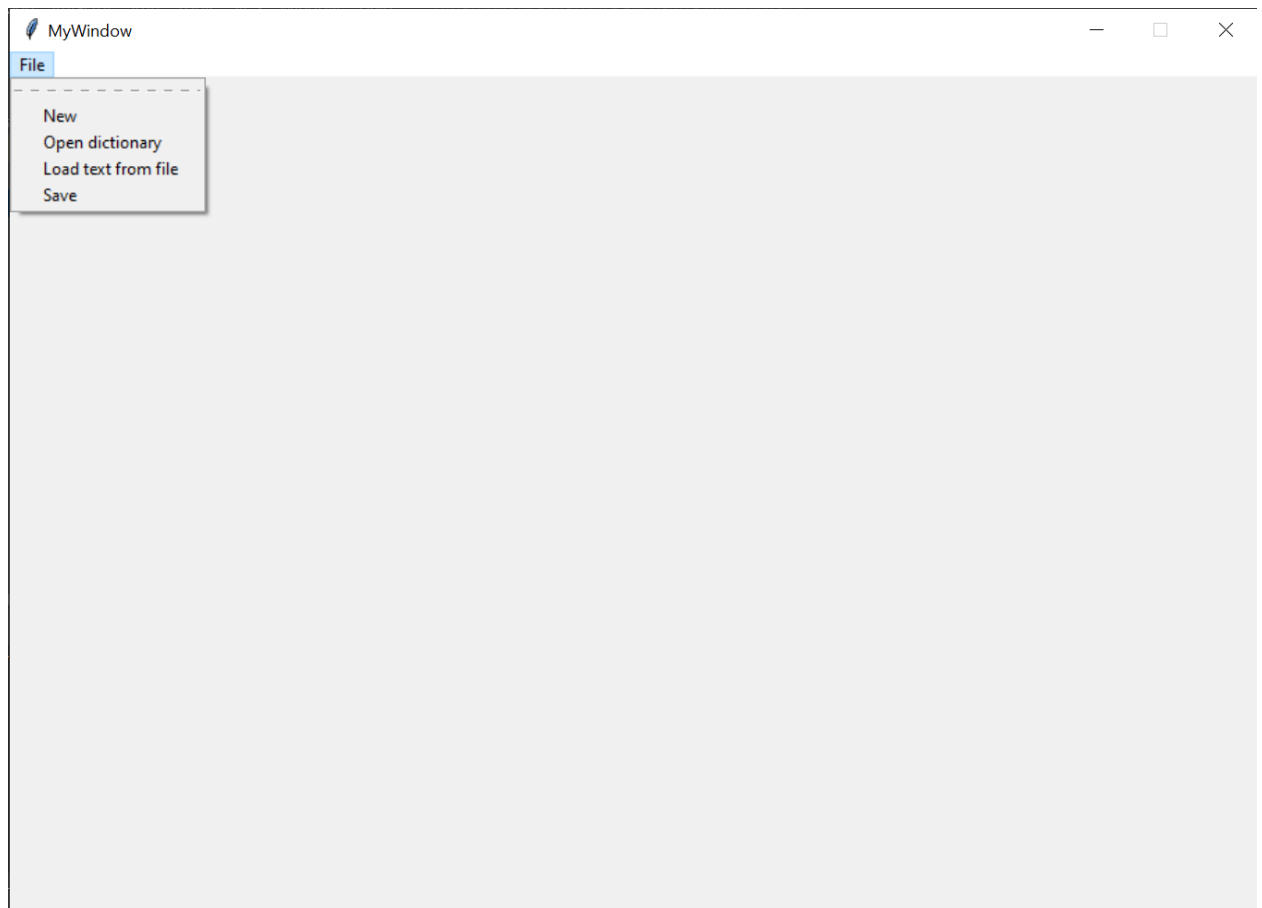
Английский язык. Список слов, упорядоченный по алфавиту и включающий как лексемы, так и словоформы, с указанием частоты встречаемости каждой из форм. Для словоформ пользователю должна быть предоставлена возможность вводить дополнительную морфологическую информацию, а именно, отнесение слова к соответствующей части речи, указание рода, числа, падежа и т.п. При этом морфологическая информация может быть оформлена как отдельная неформатированная запись, т.е. это просто текст, который пользователь может оформлять произвольным образом.

### **Ход работы**

Для выполнения лабораторной работы использовался модуль nltk, данный модуль предназначен для обработки текста, в нем имеются инструменты для разбиения текста на предложения, предложений на слова, приведения слов к начальной форме (стеминг и лемматизация).

На начальном экране приложения имеется меню, в котором пользователь может:

- Самостоятельно ввести текст (New)
- Открыть сохраненный ранее словарь (Open dictionary)
- Загрузить текст из файла
- Сохранить созданный словарь

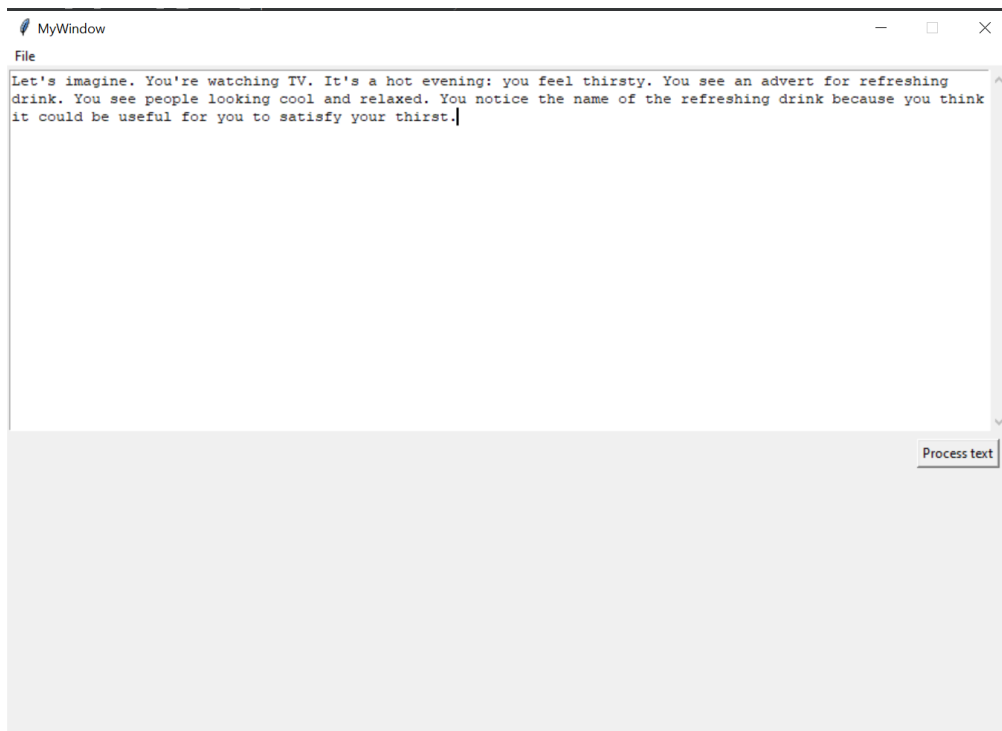


Создать новый файл (пункт меню New), где будет выведено поле ввода текста, куда пользователь может записать свой текст и обработать его



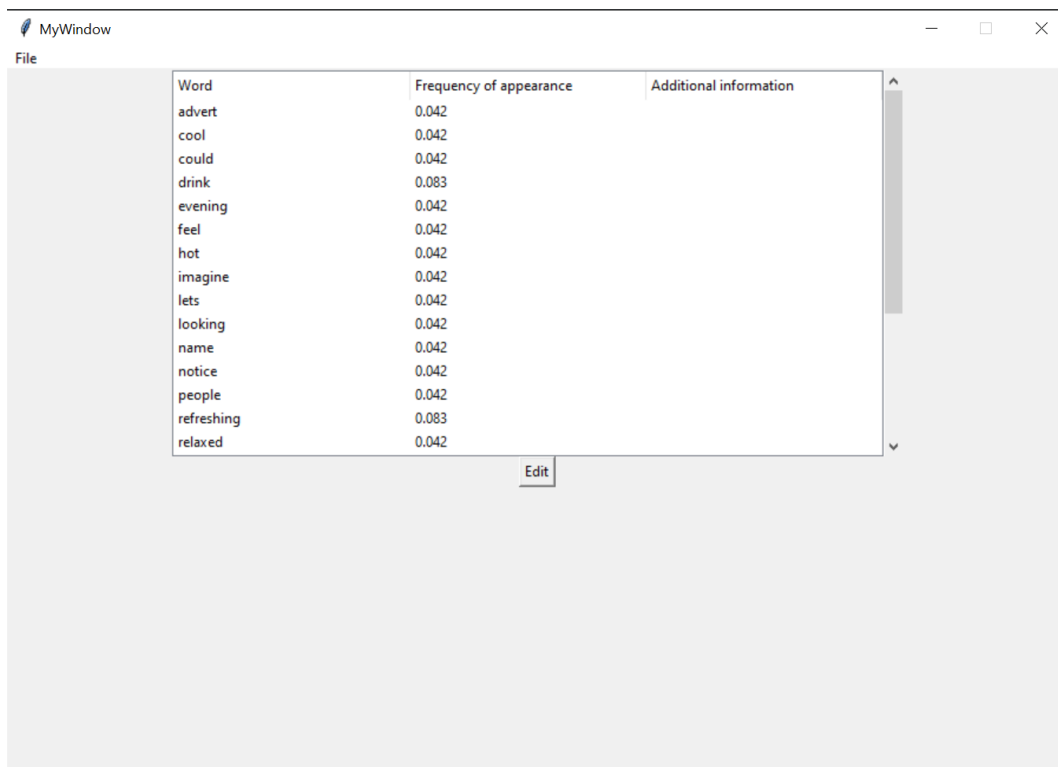
## Пользователь вводит текст

Let's imagine ... You're watching TV. It's a hot evening: You feel thirsty. You see an advert for a refreshing drink. You see people looking cool and relaxed. You notice the name of the refreshing drink because you think it could be useful for you to satisfy your thirst.

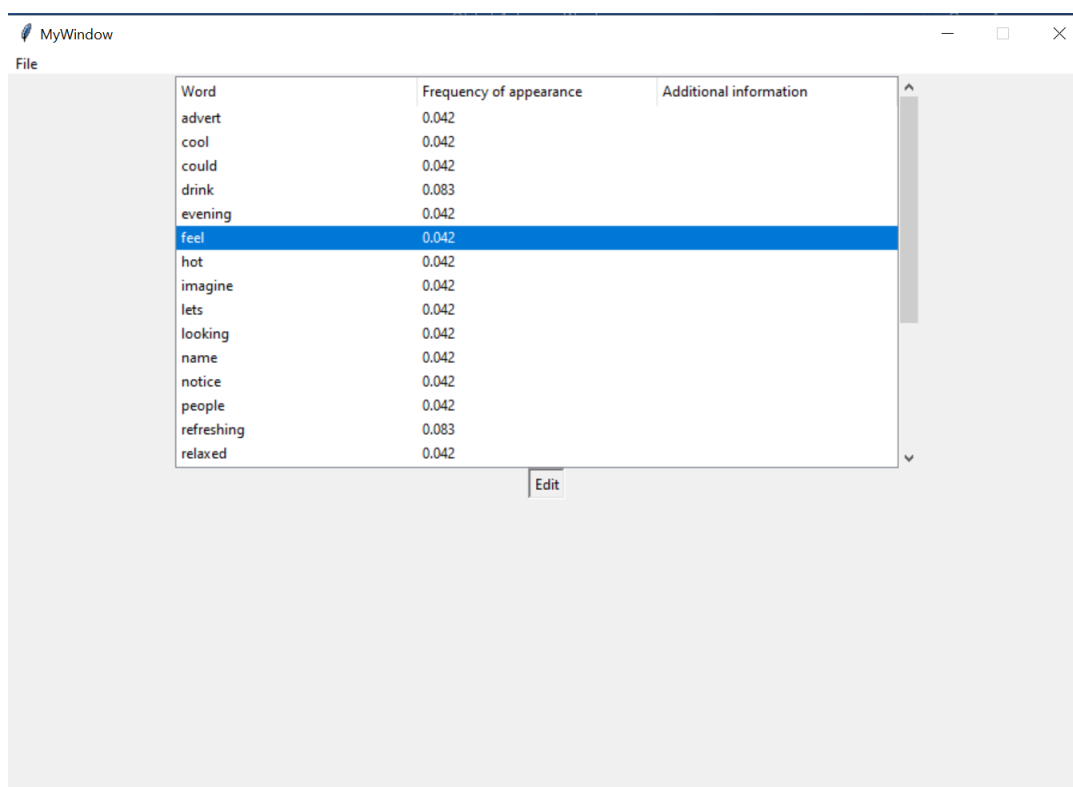


и нажимает кнопку Process text

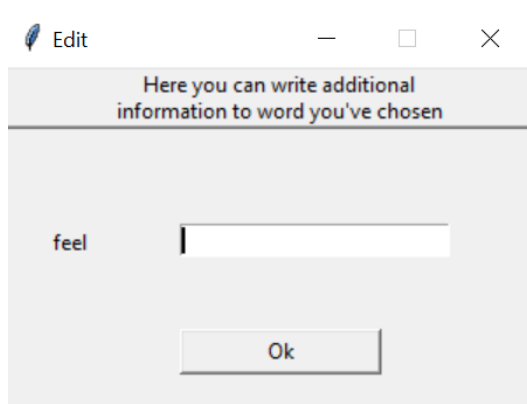
Получаем словарь, упорядоченный по алфавиту, содержащий все существенные лексемы и словоформы (артикли, междометия, местоимения отбрасываются), так же имеем графу частоты встречаемости каждого из слов в данном тексте, которая вычисляется отношением количества данного слова к количеству всех слов в данном тексте.



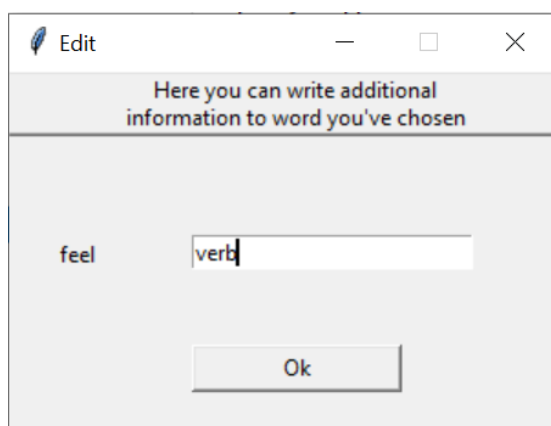
Здесь пользователь может выбрать слово и добавить информацию о нем (additional information)

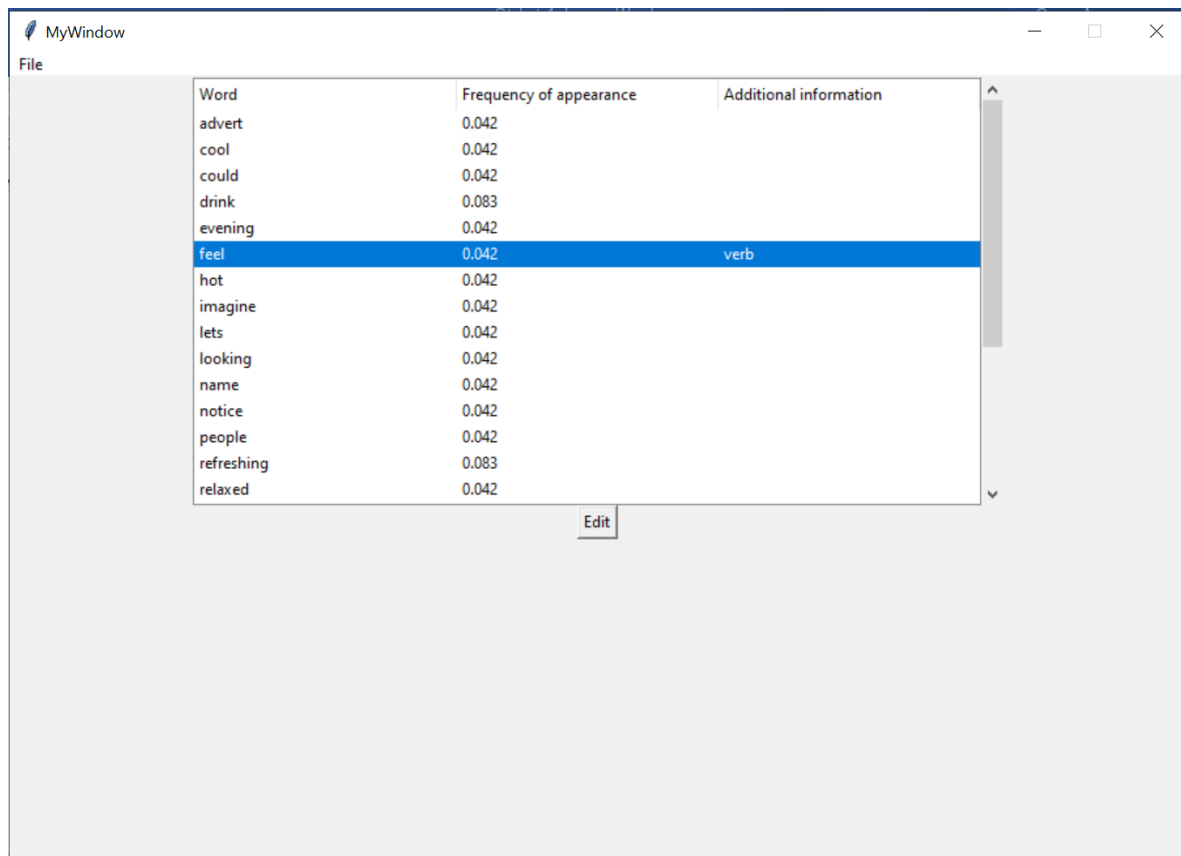


После чего появляется окно добавления информации

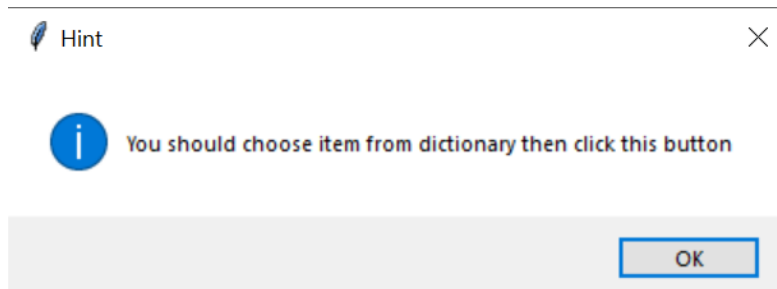


Например, добавим, что feel это глагол (verb)

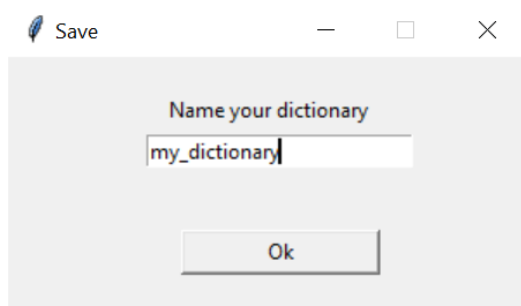
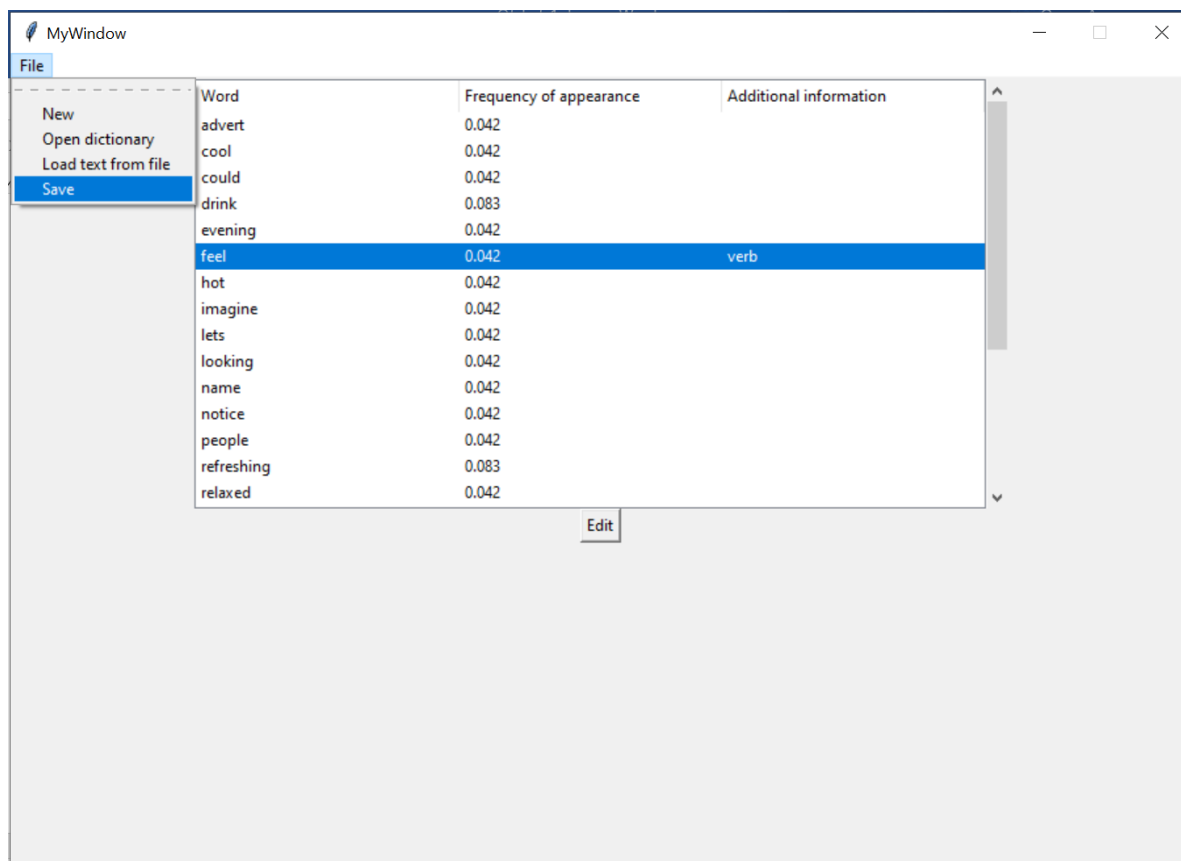




Если пользователь нажимает кнопку Edit не выбрав слово из словаря ему выводится сообщение, что он должен сперва выбрать слово, а затем нажать на кнопку.

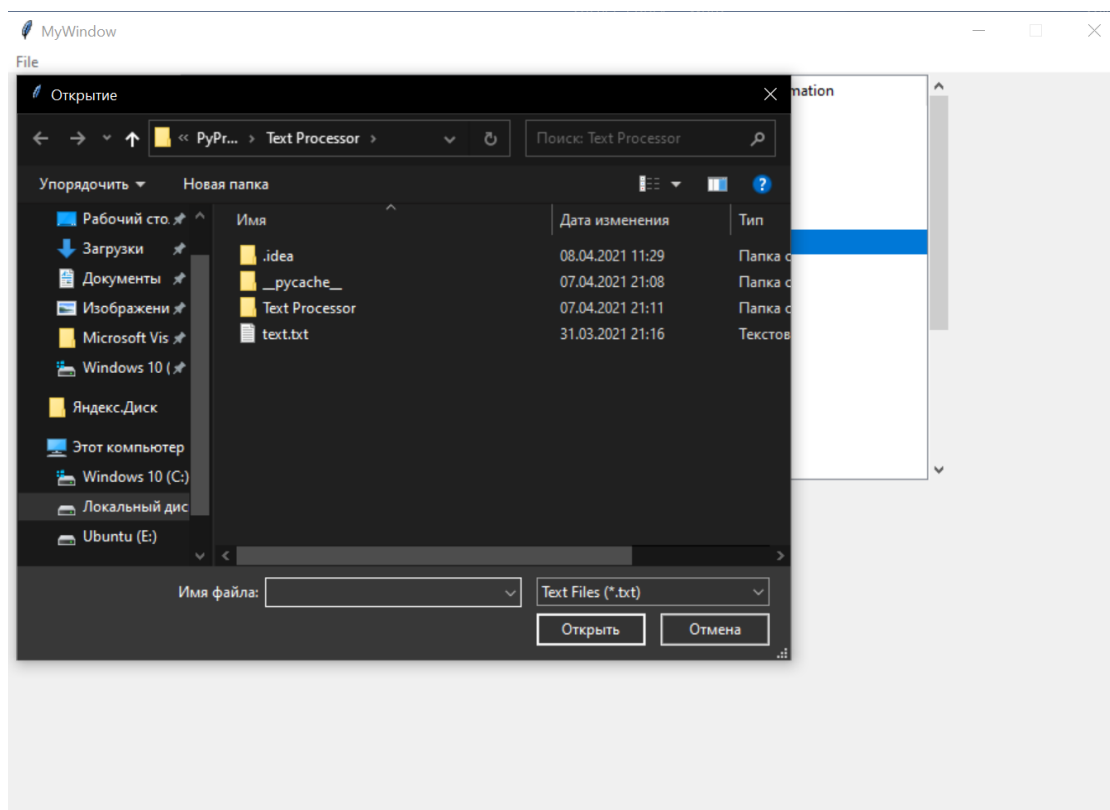
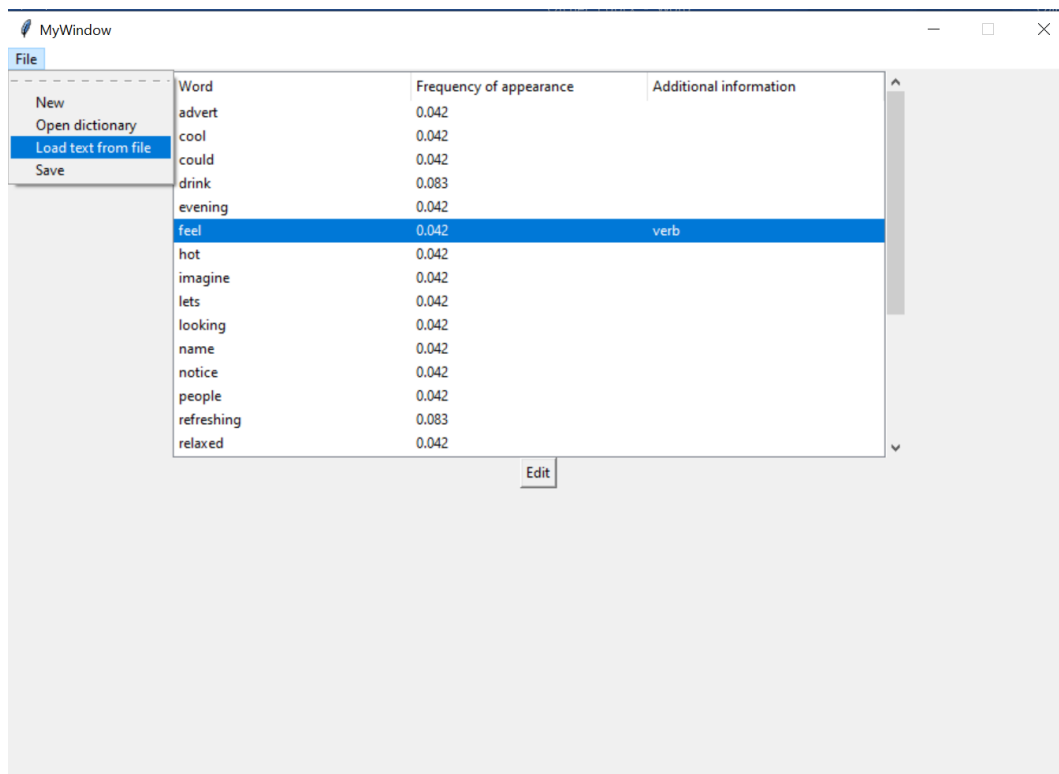


Так же можно сохранить полученный словарь в базу



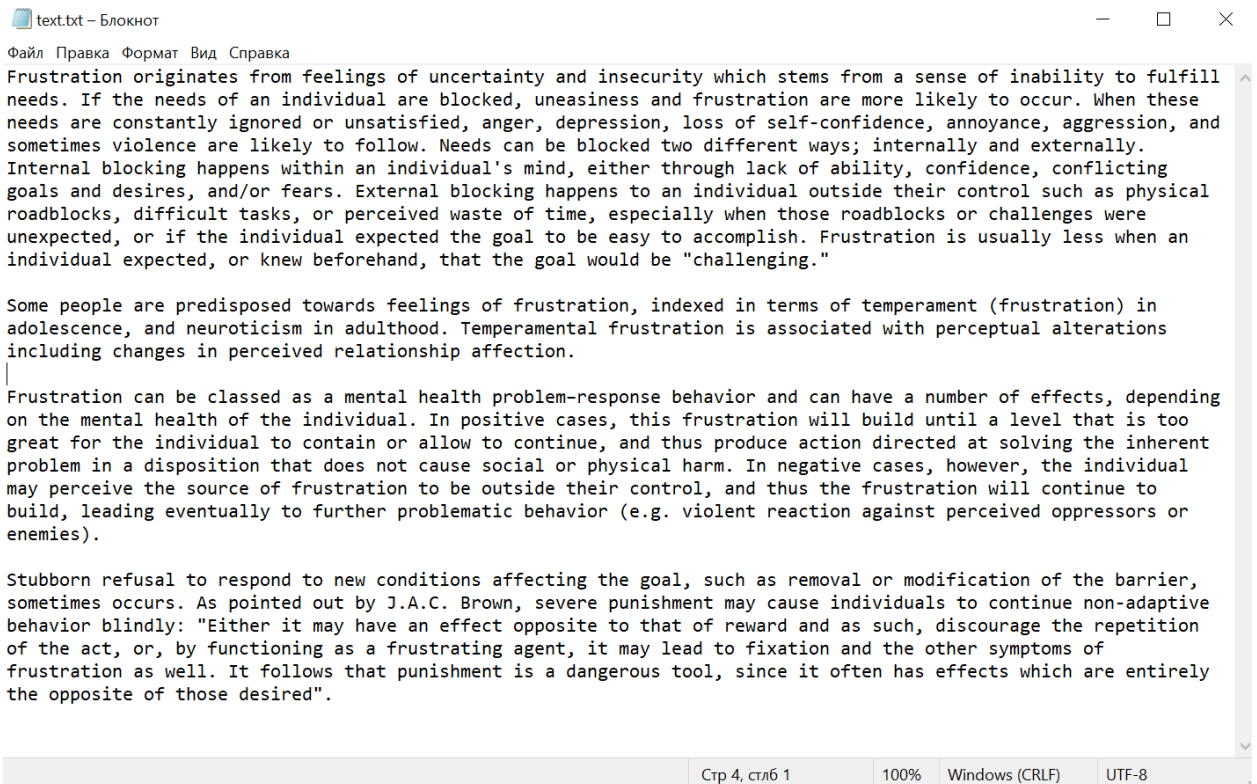
Словари сохраняются в базу данных MySQL

Так же пользователь может загрузить готовый текстовый файл

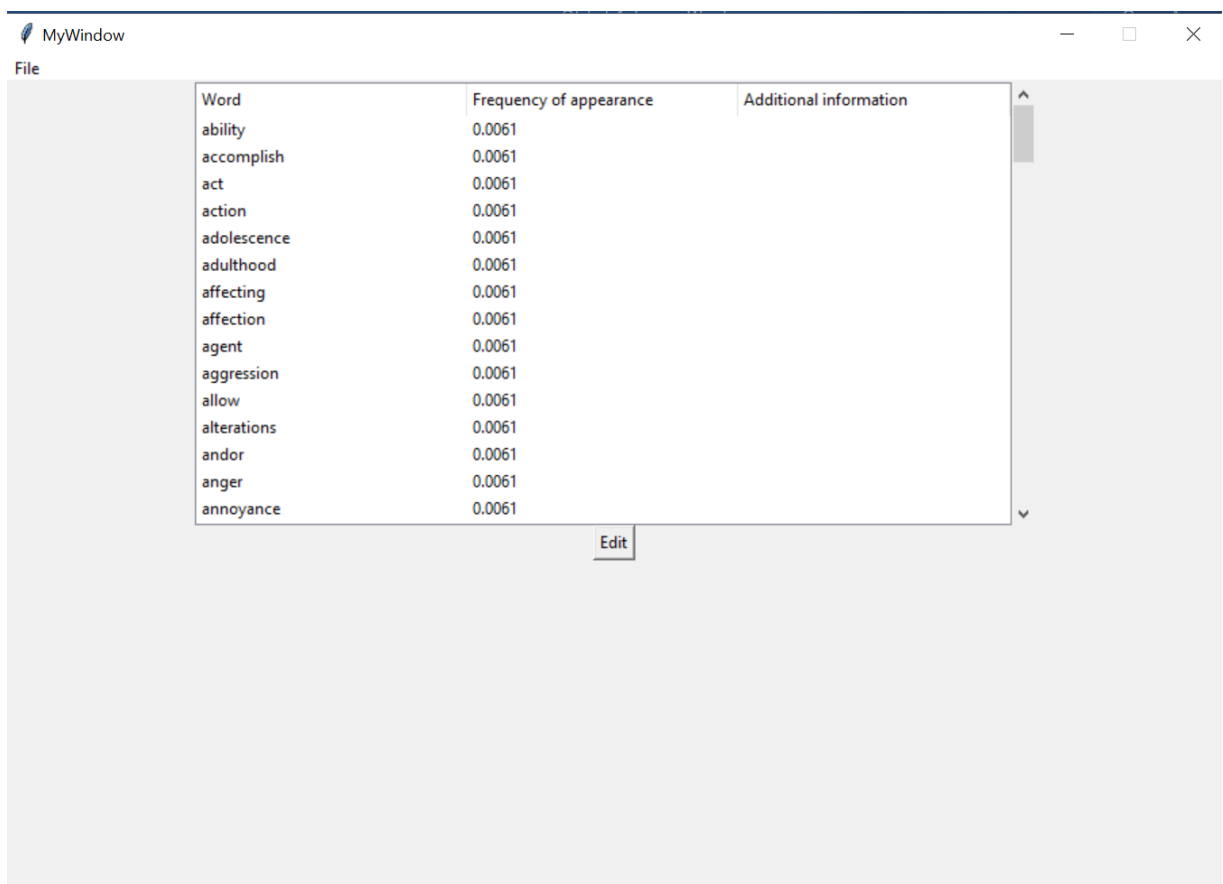


Выбираем любой текстовый файл, например у нас есть text.txt



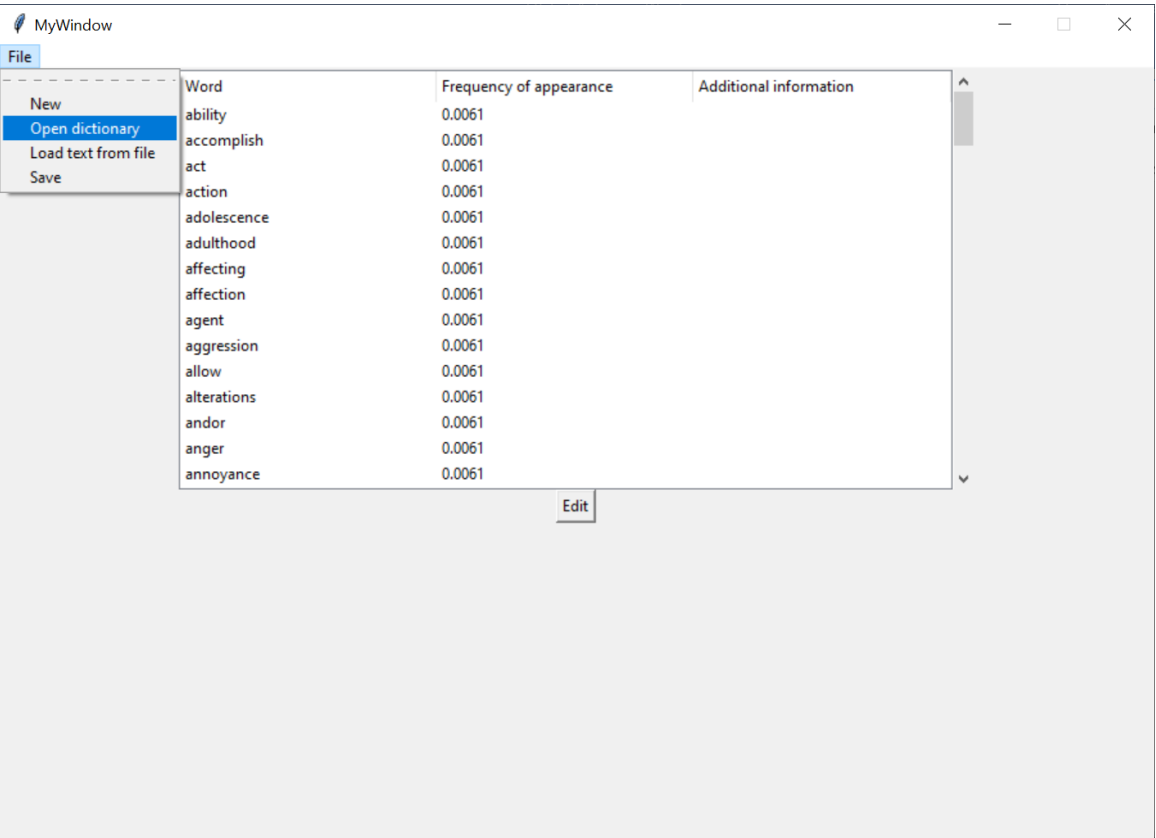


## Получаем словарь

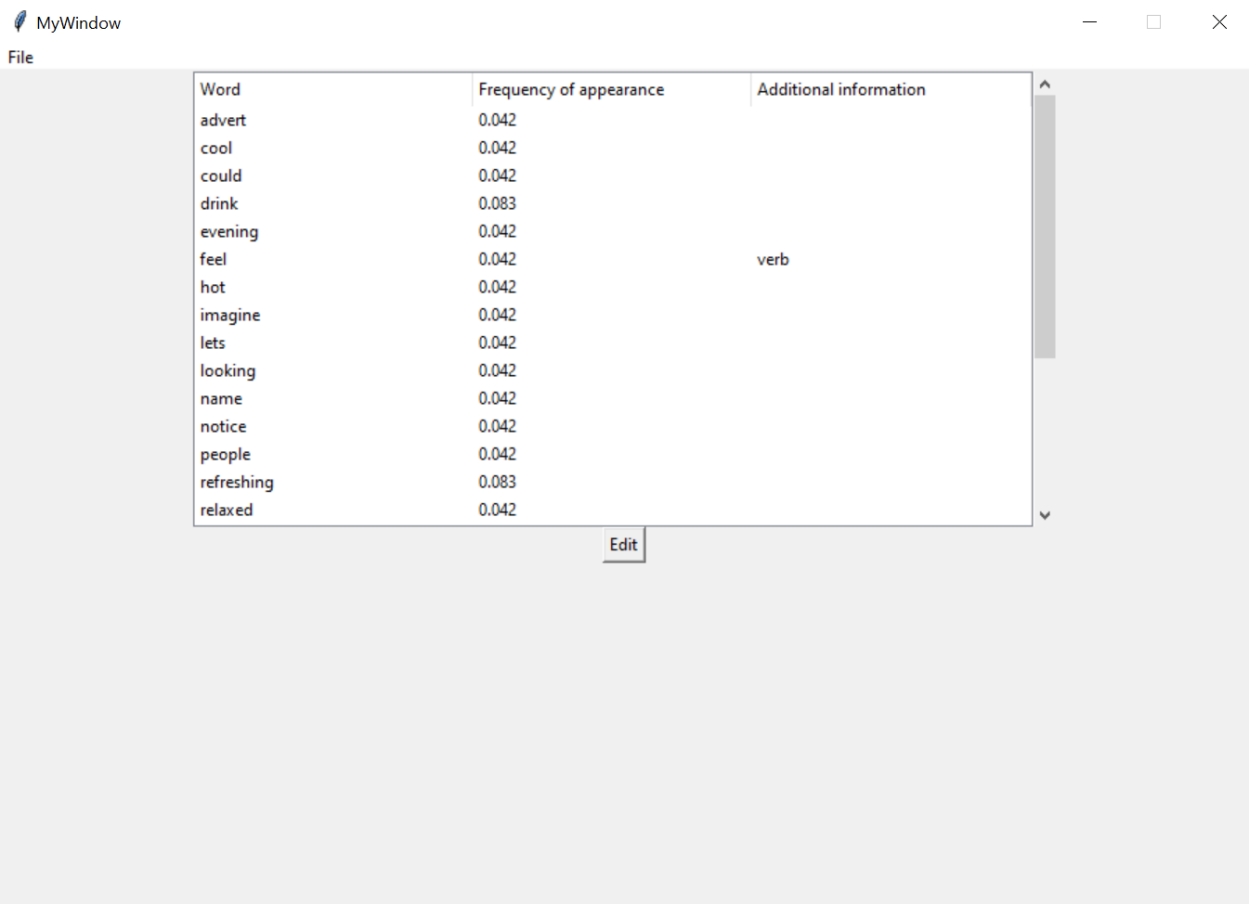


Здесь пользователь аналогично примеру с созданием своего текста может редактировать информацию о словах.

И последнее: пользователь может загрузить из базы сохраненный ранее словарь



На примере сохраненного ранее my\_dictionary



## **Вывод**

В ходе данной лабораторной работы была реализована программа создания словаря из текста. Данная программа обрабатывает текст выделяя из него существенные слова, упорядочивает их по алфавиту и вычисляет частоту встречаемости каждого слова. Имеется возможность добавления самим пользователем морфологической информации о словах. Можно создавать новый текст, загружать существующий текстовый файл, сохранять и загружать словари.