*A project report on*

**WORDZONE**

*Submitted in partial fulfillment for the award of the degree of*

**BTECH**

*by*

**RITVEAK DUGAR (16BIT0407)**



**SCHOOL OF INFORMATION TECHNOLOGY & ENGINEERING**

March,2020

**WORDZONE**

*Submitted in partial fulfillment for the award of the degree of*

**BTECH**

*by*

**RITVEAK DUGAR (16BIT0407)**



**SCHOOL OF INFORMATION TECHNOLOGY & ENGINEERING**

March,2020

**DECLARATION**

I hereby declare that the thesis entitled “WORDZONE ” submitted by me, for the award of the degree of Specify the name of the degree VIT is a record of bonafide work carried out by me under the supervision of Prof E Sathiyamoorthy.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Vellore

Date: Signature of the Candidate

**CERTIFICATE**

This is to certify that the thesis entitled “WORDZONE**”** submitted  
by RITVEAK DUGAR (16BIT0407) SCHOOL OF INFORMATION TECHNOLOGY & ENGINEERING VIT, for the award of the degree of BTECH is a record of bonafide work carried out by him under my supervision.

The contents of this report have not been submitted and will not be submitted  
either in part or in full, for the award of any other degree or diploma in this institute or  
any other institute or university. The Project report fulfils the requirements and regulations of  
VIT and in my opinion meets the necessary standards for submission.

**Signature of the Guide**  **Signature of theHoD**

**Internal Examiner**  **External Examiner**

Date: …………….…..

**CERTIFICATE BY THE EXTERNAL GUIDE**

This is to certify that the project report entitled “**WORDZONE”** submitted by R**ITVEAK DUGAR (16BIT0407)** to Vellore Institute of Technology in partial fulfilment of the requirement for the award of the degree of BTECH in Information Technology is a record of bonafide work carried out by him under my guidance. The project fulfils the requirements as per the regulations of this Institute and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

**Naveen Chandra**

**EXTERNAL SUPERVISOR**

**Product Engineer**

IDOS India Pvt Ltd

Bhive, Mahalakshmi Chambers, #29, MG Road, Bengaluru, 560001

+91 8496816918

[naveen@myidos.com](mailto:naveen@myidos.com)

**ABSTRACT**

This project is an attempt to bring in various word related paradigms in a single place, where users can learn, understand and find their words. The project has four main sections namely Find Zone, Understand Zone, Knowledge Zone and Learning Zone.

Find Zone is the section where user can find words they want, for example, rhyming words with particular meaning or a word for their crossword with particular hints or words for their scrabble game.

Understand Zone is the section where users can understand a word by looking up multiple meanings and examples of it.

Knowledge Zone is the section where users can know more related words like synonyms, antonymns, hypernymn, holonymn and hyponymn.

Learning Zone is the section which utilises the searches done in the whole application to train a machine learning model. This model smartly suggests words for the user to learn.

The whole project is aimed to improve vocabulary of users by giving them a platform to find, understand, know and learn words.

The whole project gives user a graphical user interface made on Kivy, which lets them learn and explore words with WORDZONE.

**ACKNOWLEDGEMENT**

It is my pleasure to express with deep sense of gratitude to Prof E Sathiyamoorthy, Professor Grade 1, School of Information Technology and Engineering, Vellore Institute of Technology, for his constant guidance, continual encouragement, understanding; more than all, he taught me patience in my endeavour. My association with him is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of <area>.

I would like to express my gratitude to Dr.G.Vishwanathan, Mr. Sankar Viswanathan, Dr. Anand A. Samuel, Dr. S. Narayanan, and Dr Balakrushna Tripathy, School of Information Technology and Engineering, for providing with an environment to work in and for his inspiration during the tenure of the course.

In jubilant mood I express ingeniously my whole-hearted thanks to <Program char-name>. <Program Chair and designation>, all teaching staff and members working as limbs of our university for their not-self-centred enthusiasm coupled with timely encouragements showered on me with zeal, which prompted the acquirement of the requisite knowledge to finalize my course study successfully. I would like to thank my parents for their support.

It is indeed a pleasure to thank my friends who persuaded and encouraged me to take up and complete this task. At last but not least, I express my gratitude and appreciation to all those who have helped me directly or indirectly toward the successful completion of this project.

Place: Vellore

Date: **Ritveak Dugar**

**CONTENTS**

**CHAPTER 1 - INTRODUCTION**

* 1. FIND ZONE 9

1.2 UNDERSTAND ZONE 9

1.3 KNOWLEDGE ZONE 9

1.4 LEARNING ZONE 10

**CHAPTER 2 – LITERATURE SURVEY 11**

**CHAPTER 3 – DESIGN**

3.1 UML DESIGN 13

3.1.1 FIND ZONE 13

3.1.2 KNOWLEDGE ZONE 14

3.1.3 UNDERSTAND ZONE 15

3.1.3 LEARNING ZONE 16

3.2 SYSTEM ARCHITECTURE 17

**CHAPTER 4 – SOFTWARE REQUIREMENTS**

4.1 PROGRAMMING LANGUAGE 18

4.1.1 PYTHON 18

4.1.2 KIVY 18

4.2 LIBRARIES 19

4.2.1 NLTK 20

4.2.2 SPELL CHECKER 21

4.2.3 PRONOUNCING 21

4.2.4 DATAMUSE AND OXFORD API 24

**CHAPTER 5 – MODULE DESCRIPTION**

5.1 FIND ZONE 25

5.1.1 RHYMING ZONE 25

5.1.2 CROSSWORD HELPER 30

5.1.3 SCRABBLE HELPER 35

5.2 UNDERSTAND ZONE 40

5.3 KNOWLEDGE ZONE 43

5.3.1 SIMILAR MEANING WORDS 43

5.3.1 OPPOSITE MEANING WORDS 43

5.3.1 HYPERNYM / HYPONYMN / HOLONYM 43

5.4 LEARNING ZONE 50

**Chapter 1**

**Introduction**

WORDZONE is an umbrella that covers a lot of word related operations. A user can find words, understand them, get to know more related words and learn new ones which are suggested based on their searches.

Therefore giving a wide range of functionalities to the user to play around words and increase their knowledge as well.

Wordzone has four main sections, which combined makes wordzone a one of its own platform:

1. Find Zone
2. Understand Zone
3. Knowledge Zone
4. Learning Zone

1.1 FIND ZONE

FIND ZONE allows the user to find words that they are looking for.

A user can look up for words which rhyme with a given word, further the result can be narrowed down to be of some specific meaning in the “Rhyming Words” section.

User can also look up for words with some meaning, which can be narrowed down to a list of words with fixed character at a position in the “Crossword Helper” section.

Words for Scrabble can be searched using “Scrabble Helper” where the word fragment, its position and length of the word can be found.

* 1. UNDERSTAND ZONE

UNDERSTAND ZONE allows user to find the meaning and examples of the word they enter. It also shows multiple meanings that a word might have and returns example for each case as well.

* 1. KNOWLEDGE ZONE

KNOWLEDGE ZONE allows user to know more words around the word they enter. A user can find Synonyms in the “Similar meaning” section, antonyms in “Opposite meaning” section and other related words in “Hypernymn/Hyponymn/Holonymn” section

* 1. LEARNING ZONE

LEARNING ZONE is the section that makes use of all the searches made by the user in the application to smartly suggest new words to learn. The model retrains every time the learner learns a new word or selects “I already know this word” option. Hence making WORDZONE an evolving model which not only gives user a platform to find, understand and know words, but makes their learning personalised as well.

**Chapter 2**

**Literature Survey**

A lot of projects have been made previously on NLP using the features, tools and techniques that are used in wordzone.  
Below is a literature survey on few of those tools, techniques and properties.

|  |  |  |  |
| --- | --- | --- | --- |
| **Ref no.** | **Topic** | **Summary** | **Use/Meaning** |
| 1. | Web Corpus[1] | Traditionally written corpora are primarily recorded from print media, with advent of internet, web data can be used to train corpora as well. | Corpora are basically huge collections of words and their associated features like meanings, pronunciations. They are used for NLP based usages |
| 2. | NLTK[2] | The need of an umbrella covering a lot of linguistics related paradigm gave birth to NLTK toolkit. | NLTK being a collection of modules which helps in computation of linguistics is extensively used in projects related to NLP. |
| 3. | WordNet[3] | WordNet is a lexical database for the English language. It is basically a combination of dictionary and thesaurus which contains words, their definitions and many other related words. | Wordnet acts as database which is used for fetching properties of words and their related words. |
| 4. | Semantic Analysis[4] | Latent Semantic Analysis is a good approach for finding accurateresults. Another important use of LSA is to find out semantic similarities between different set of textual data. | Semantic Analysis helps in determining the meaning of sentences. |
| 5. | Phonetics[5] | Phonology is used to recognize sound so as to find out the language, semantic and syntactic meaning. It also helps in pronouncing words and finding phonetically similar words. | Stores data related to pronunciation of words, hence used in projects related to rhyming words. |
| 6. | Python GUI [6] | There are many GUI libraries forPython programs. The most famous ones are: Kivy,Tkinter, PyQt, PyGUI,Pyforms, PySide, Flexx, PySimpleGUI, IPyWidgets, Wax Python GUI, etc. | To give a Graphical User Interface to the python programs. |
| 7. | Dictionary[7][8] | Dictionary requires corpus which contains words and its attributes like meaning, pronunciation, example etc. so that it can also be used for machine translation and language processing. | In addition to using dictionary for finding meanings, one can also use it to find rhyming words and pronunciations. |

**Chapter 3**

**Design**

Design is an integral part of development, it paves the pathway for the development.

Below are the designs made for wordzone.

3.1 UML Diagrams

3.1.1 Find Zone:



3.1.2 Understand Zone:

****

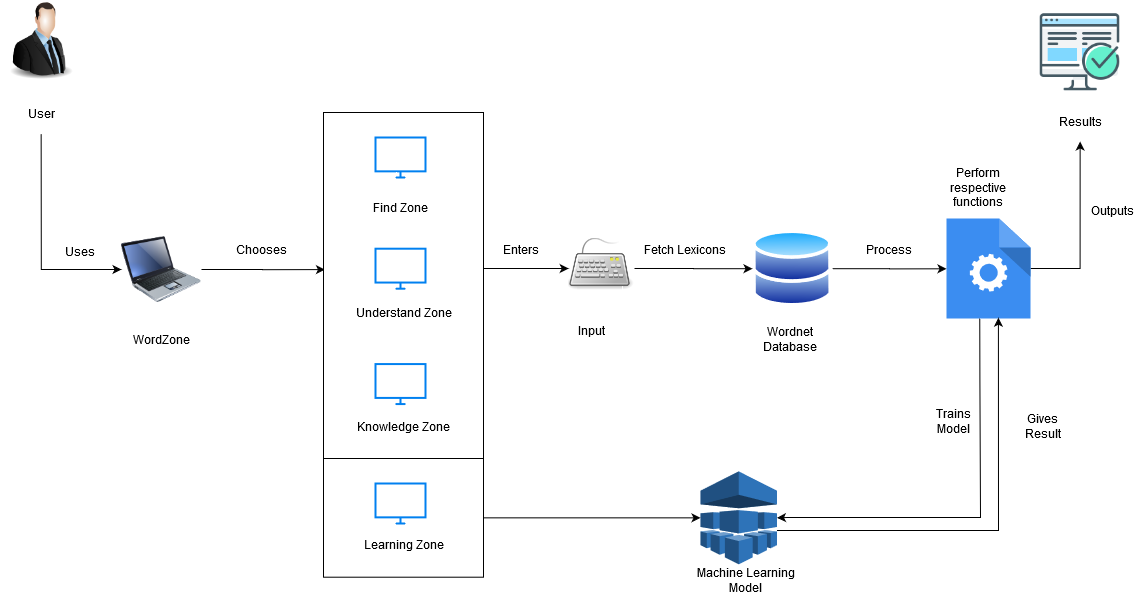
3.1.3 Knowledge Zone:

****

3.1.4 Learning Zone:

****

3.2 System Architecture

****

**Chapter 4**

**Software Requirements**

4.1 PROGRAMMING LANGUAGES

Programming language is the base for any software being developed; they act as the field where the crop grows. Hence there is always a particular kind of field which is best for a given crop.

The following are the programming languages that are used in WORDZONE.

4.1.1 PYTHON

Python is a high level programming language which is always preferred over others when it comes to Machine Learning and Natural Language Processing, because of the wide variety of libraries that it supports.

Since WORDZONE uses NLP and ML, python was the go to language.

Python being easy and widely used has a lot of support online, hence development and troubleshooting is eased because of the rich documentation and forums online.

Many libraries like NLTK, pronouncing, spellchecker allowed implementation and raw data for processing in the software.

The function written in python then processes the data and returns result.

Python in WORDZONE is primarily used for the all round processing, and hence it can truly be called as the soul of wordzone.

This soul however needed a body as well, so to provide a Graphical User Interface, Kivy was selected.

4.1.2 KIVY

Kivy is an open source python framework for developing user interfaces. There were other options for GUI like Tkinter, Pyqt and many others. Kivy stood out in the line because the modern look and the fact that it is more customisable.

Kivy has its own syntax for defining screens, buttons and all other graphical components. Thereby aiding in development of rich but simplistic User Interface.

4.2 TOOLS AND LIBRARIES

There are few tools and libraries which helped in building WORDZONE apart from the basic programming languages. All these tools and libraries are freely available and have good documentation to understand and work with them.

4.2.1 NLTK

NLTK – Natural Language Toolkit is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language. This is one of the most usable and mother of all NLP libraries.

It consists of the most common algorithms such as tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition. NLTK helps the computer to analysis, pre-process, and understand the written text.

Since wordzone only plays around words and not the syntactic meanings, wordnet is used at most. Wordnet is a lexical database which records semantic relation of words including synonyms, hyponyms, and meronyms. The synonyms are grouped also known as synsets with short definitions and usage examples. This comes in handy for making a rich group of related words.

Wordnet acts as a corpus, which is nothing but a collection of words. In the view of wordzone, wordnet acts as our database which has a lot of words along with some associated meaning and properties. The words are in turn linked to each other forming a network of words which helps to find related words to a given word.

The logic whatsoever required is applied on data collected from wordnet which can be made rich by including more and more related words.

For example in Rhyming words section, where the rhyming word is required to be related to something, a rich collection of words are made by using wordnet:

syns=wordnet.synsets(mword)

        for syn in syns:

            rr+=syn.lemma\_names()

        #Since simple synonym set is not enough, lets add more related words

        #finding all related words among which rhyming words is to be found

        hr=[]

        syns=wordnet.synsets(mword)

        for syn in syns:

            sn=syn.hypernyms()#broader category:colour is a hypernym of red.

            an=syn.hyponyms() #narrower category - red : color

            dn=syn.member\_holonyms()#Body is a holonym of arm, leg and heart

            for s in sn:

                hr+=s.lemma\_names()

            for a in an:

                hr+=a.lemma\_names()

            for d in dn:

                hr+=d.lemma\_names()

        #now even "loaf" gets included when "food" is given as input

        #making the list richer by adding synonyms of the words which are in hr.

        fn=[]

        for h in hr:

            ss=wordnet.synsets(h)

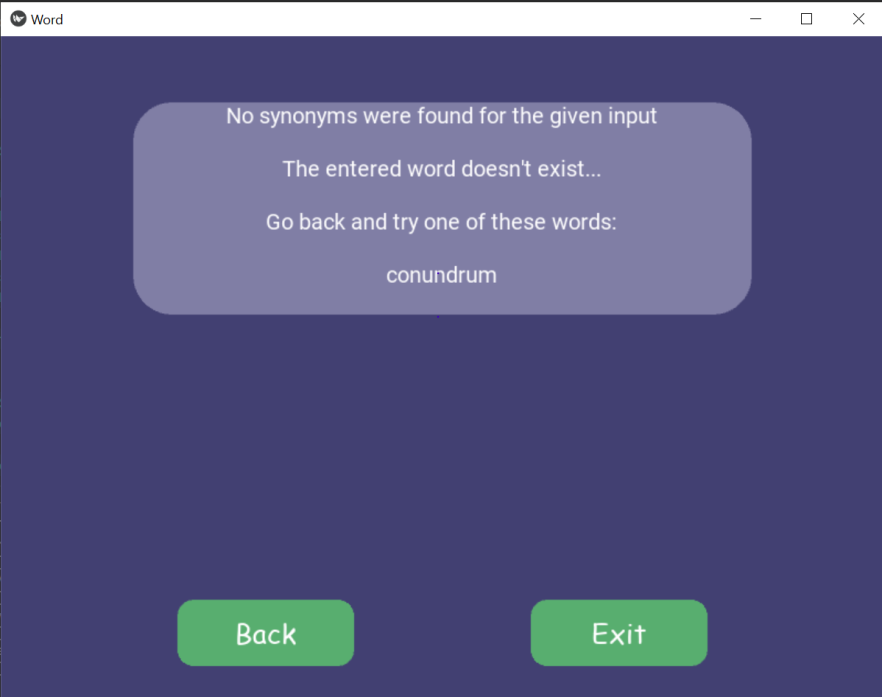
            for s in ss:

                fn+=s.lemma\_names()

Thereby making the collection rich and then applying logic.

4.2.2 SPELL CHECKER

Spell checker is a library in Python which suggest words near the wrongly spelled words. This helps in the result pages when there is no result due to wrong spelling.

Below is an example when the entered word was “Conundrom” instead of “Conundrum”:

Code snippet:

        if(len(synonyms)>0):

            ss=set(synonyms)

            new = ""

            for x in ss:

                new += x+"\n"

            return new

        else:

            str="No synonyms were found for the given input\n\n"

            if(spell.correction(word)!=word):

                str+="The entered word doesn't exist...\n\n"

                if len(spell.candidates(word))>0 :

                    str+="Go back and try one of these words:\n\n"

                for s in spell.candidates(word):

                    str+=s+"\n"

            return str

4.2.2 PRONOUNCING

Pronouncing is a library in python which helps in finding similar sounding words, it makes use of CMU Pronouncing Dictionary which holds the phonetics of words.

Code Snippet:

def rhy(self,word):

        if(word==""):

            return "No word entered"

        fb = pronouncing.rhymes(word)

        if(len(fb)>0):

            ss=set(fb)

            new = ""

            for x in ss:

                new += x+"\n"

                #add more details like pos and all

            #print(new)

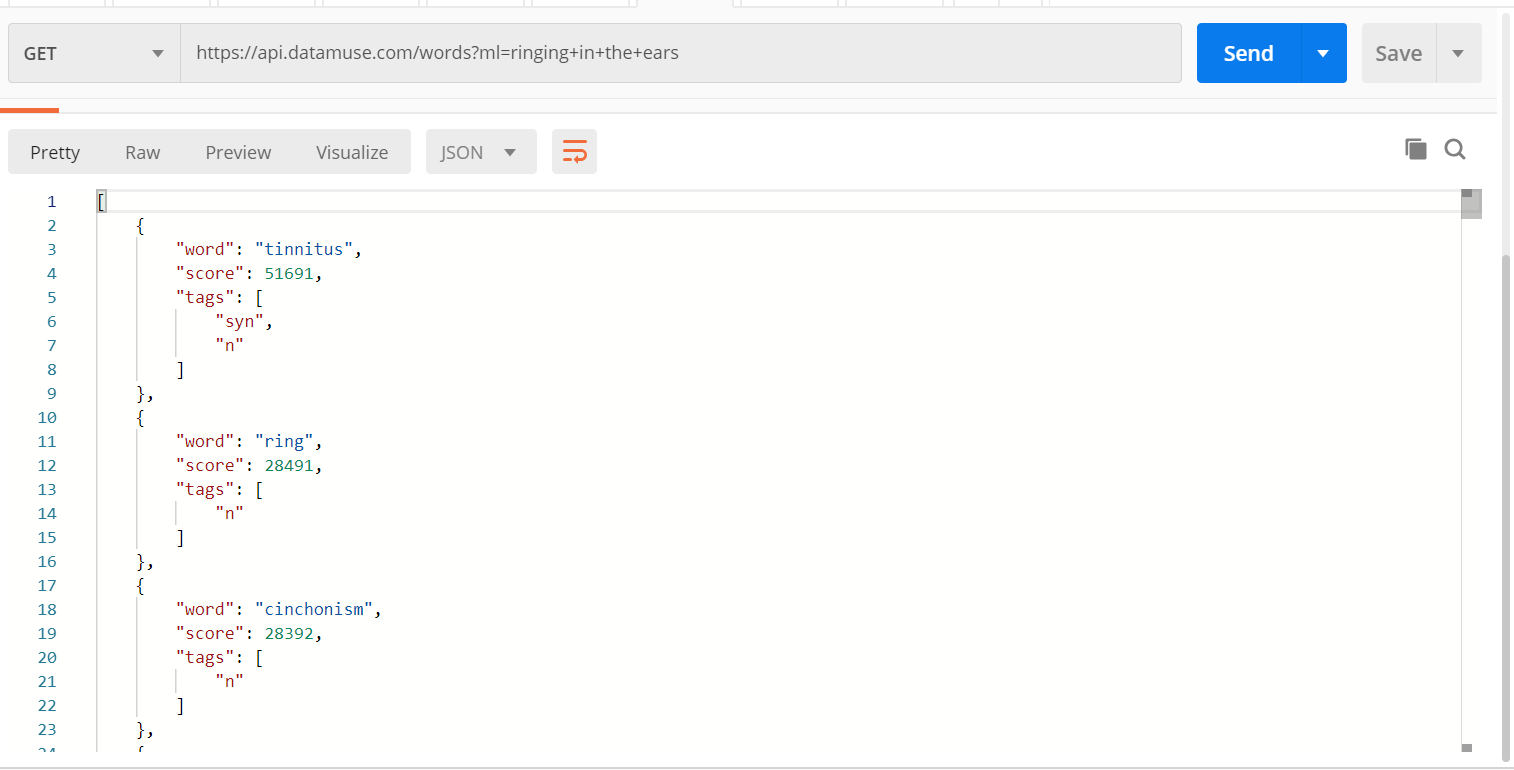
            return new

4.2.4 DATAMUSE AND OXFORD API

The Datamuse API is a word-finding query engine for developers.

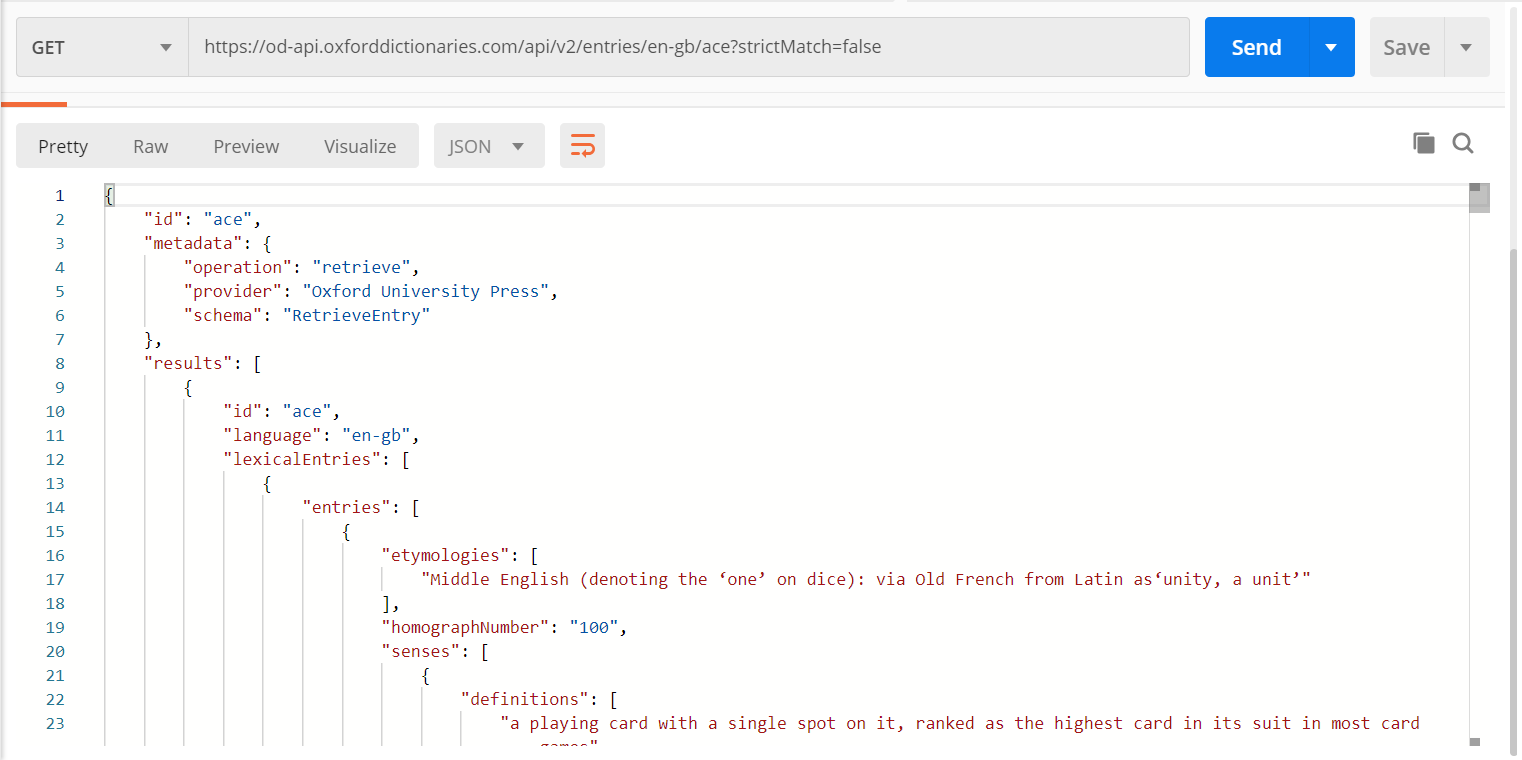
One can use it in their apps to find words that match a given set of constraints.

For example searching for words related to “Ringing in ears” in Crossword Helper:



Oxford Dictionary API uses official oxford dictionary for returning word meanings.

For example, it returns all the possible meanings with examples:



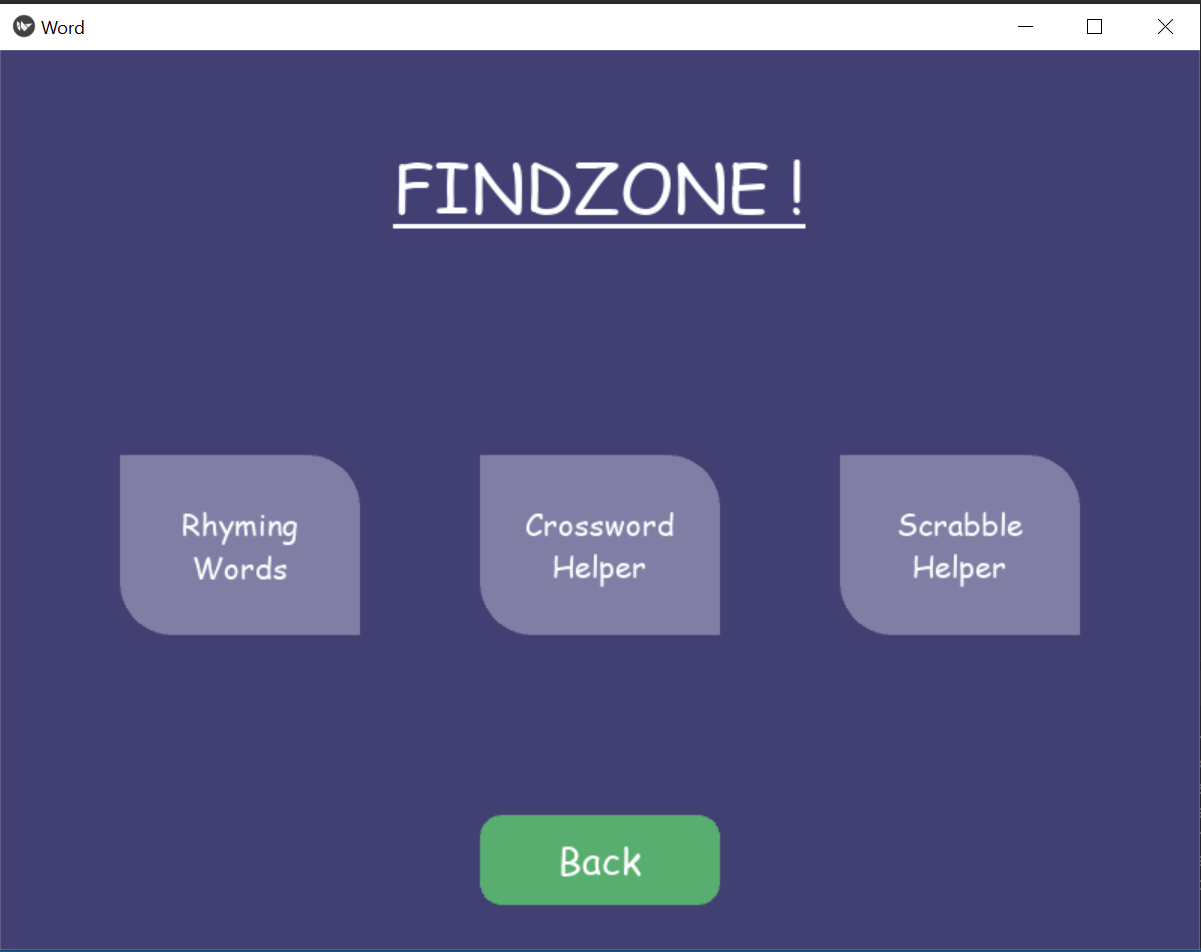
**Chapter 5**

**Module Description**

WORDZONE has four main modules namely - FIND ZONE, UNDERSTAND ZONE, KNOWLEDGE ZONE and LEARNING ZONE. All these modules further have different sections. These sections together contribute to train a model which helps in personalised word suggesting model which helps the user to learn new words.

5.1 FIND ZONE

Find Zone comprises of sections which helps the user in finding words with constraints which belong to either of the three sections:

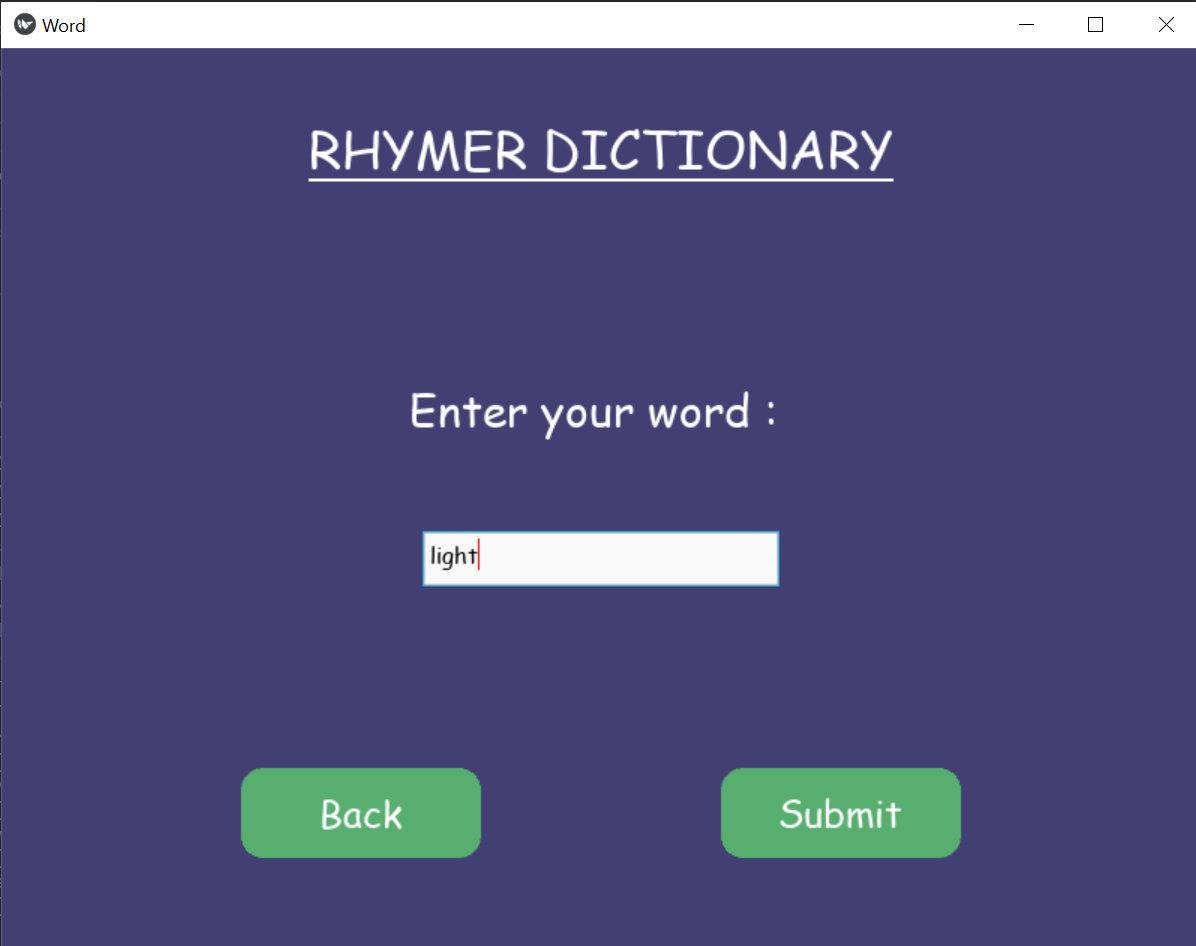


5.1.1 RHYMING ZONE

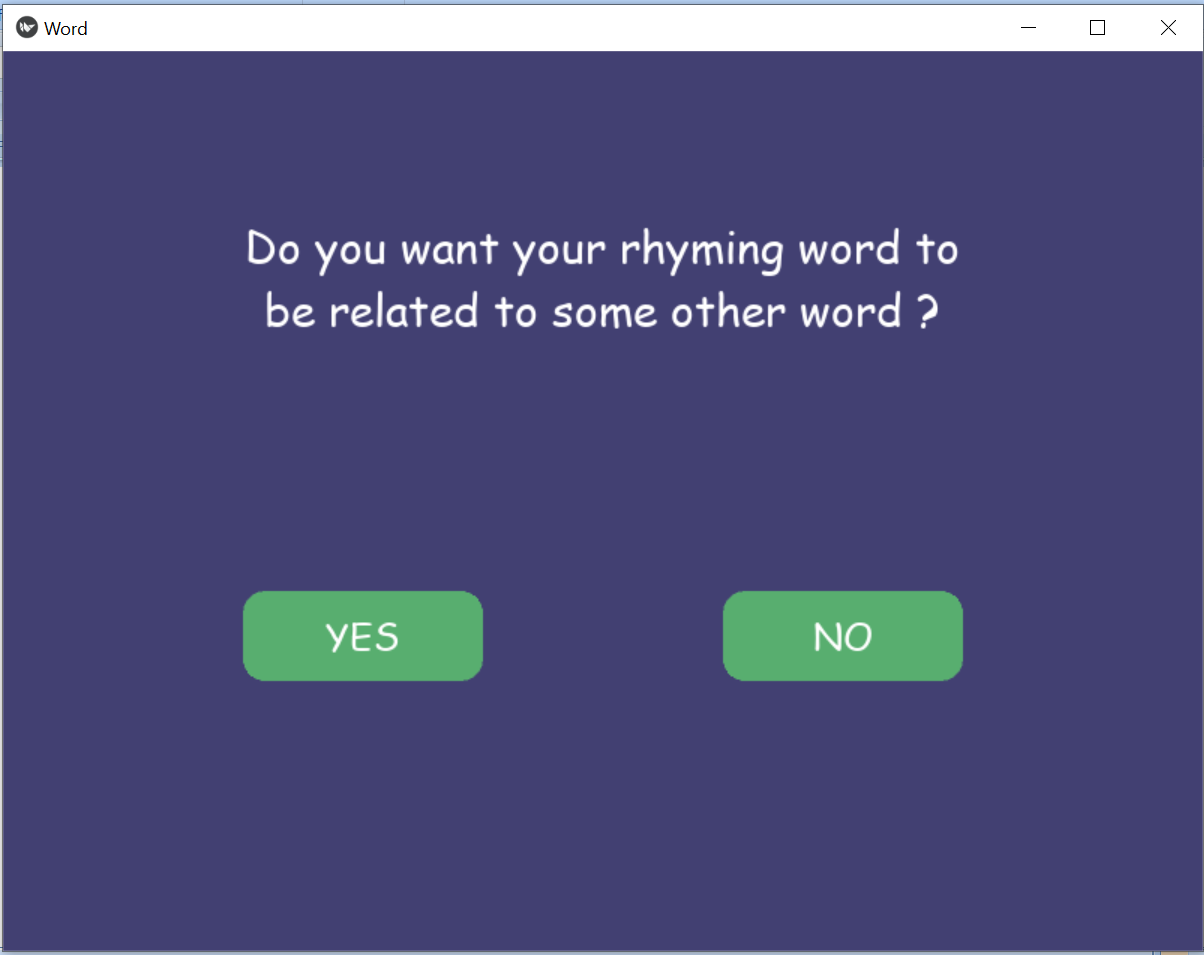
Who doesn’t like to put a bit of rhyme into their writings once in a while! This is the section where user can find words rhyming to a particular word, additionally; the user can input a word with which the rhyming word should be related to, thereby easing the way for poets who are looking for a rhyming word with a particular word but with a particular meaning.

Snippets with the workflow

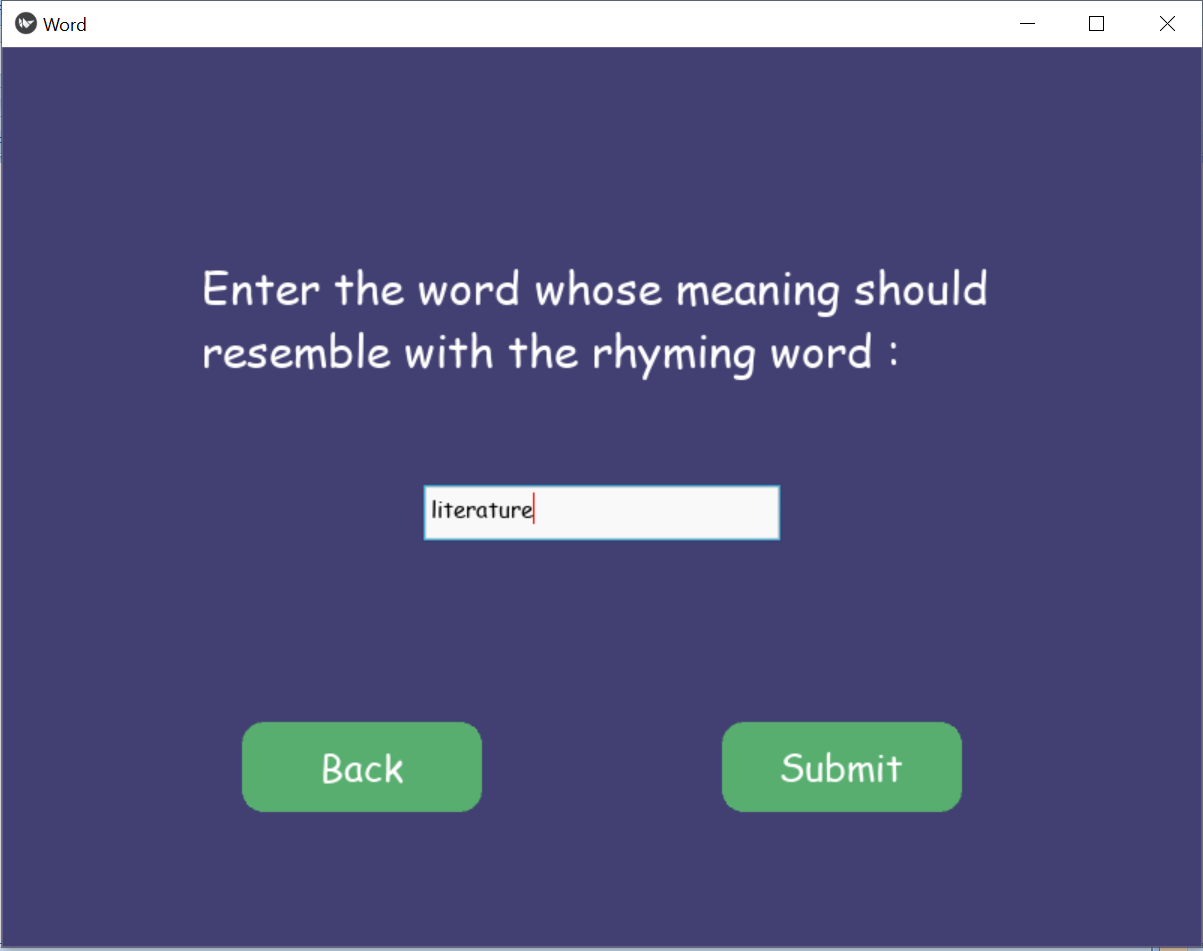
Taking Input:



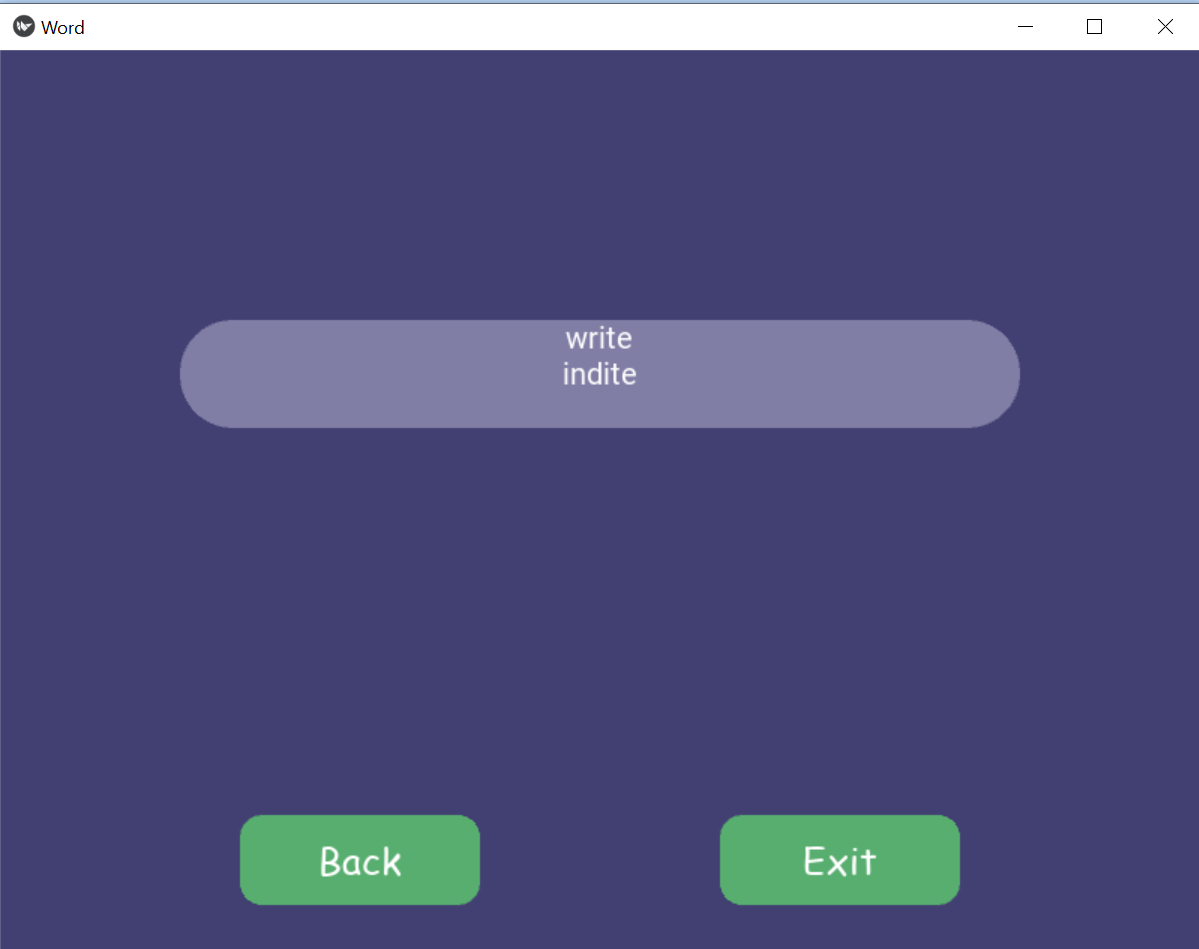
Choosing if the word should be of particular meaning:



Entering the word, the results should be related to:



The results:



5.1.2 CROSSWORD HELPER