****

**School of Information Technology and Engineering**

**Winter Semester 2019-20**

**Capstone Project Review II**

**Title – WORDZONE**

**Submitted by :**

**Ritveak Dugar**

**16BIT0407**

**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.

This project uses NLP to fulfill these. The following are implemented in the project:

1. Findzone - Where user can find words like :
   1. Find Similar words
   2. Find Opposite words
   3. Find Rhyming words (with and without a particular meaning)
   4. Crossword Helper
2. UnderstandZone – Where user can understand their word.
3. KnowledgeZone – Where user can know more words related to the entered word.

To give user a good interface, I will be using Kivy for GUI.

I have made a UI Wireframe using Adobe XD which can be accessed [here](https://xd.adobe.com/view/d214429e-8809-458a-6d4b-6627d385a3fe-a425/).

(The wireframe doesn’t accept inputs, but the buttons redirects to respective screens)

**Required Stack**

1. Python

2. NLTK

3. Kivy – For GUI

**Literature Survey**

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

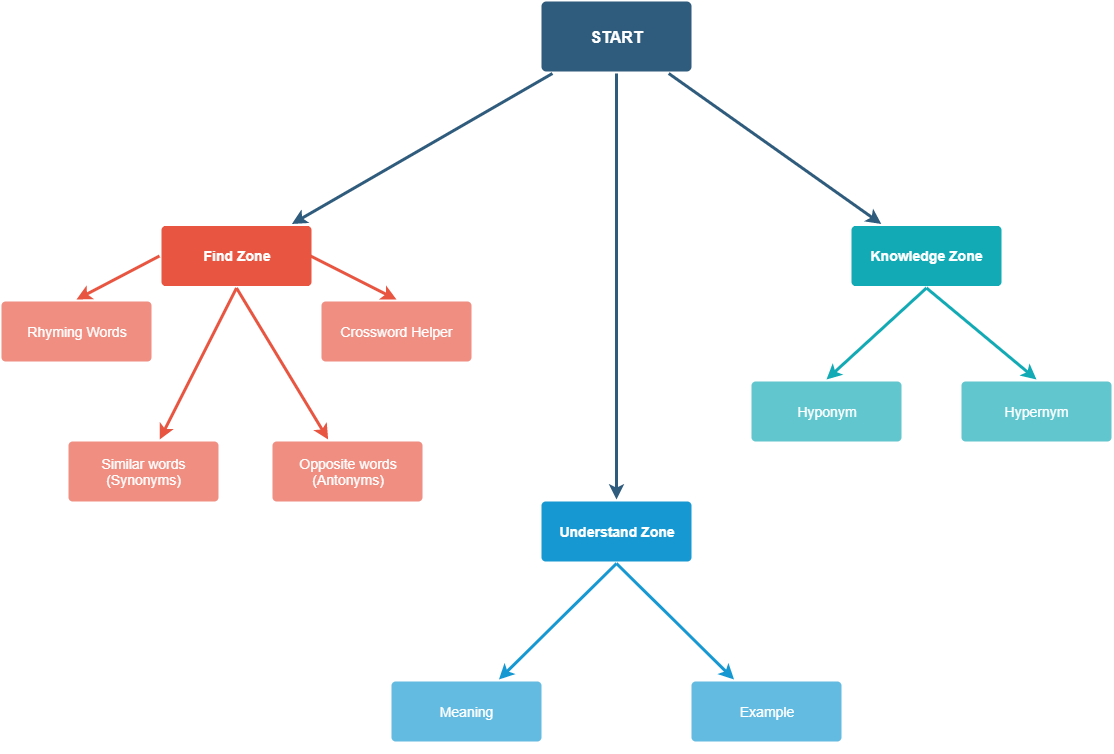
|  |  |  |  |
| --- | --- | --- | --- |
| **Ref no.** | **Topic** | **Summary** | **Use/Meaning** |
| 1. | Web Corpus[1] | Traditionally written corpora are primarily recorded from print media, with advent of internet, web data can be used to train corpora as well. | Corpora are basically huge collections 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. |

**Design**

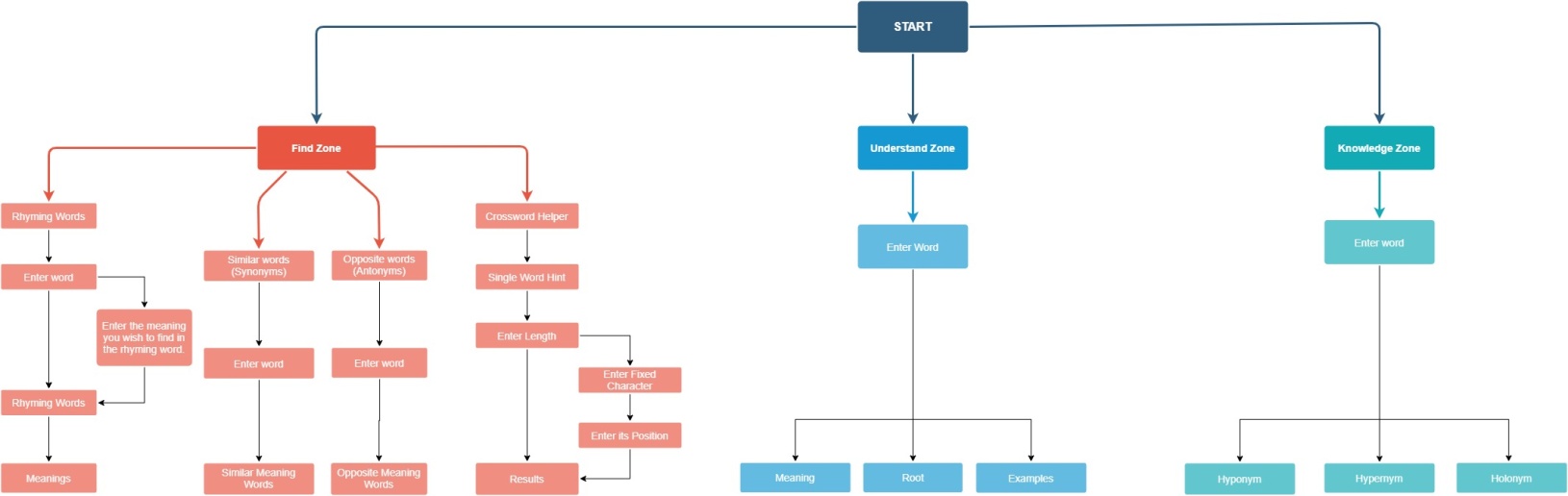
Wireframe hosted at:

<https://xd.adobe.com/view/d214429e-8809-458a-6d4b-6627d385a3fe-a425/>

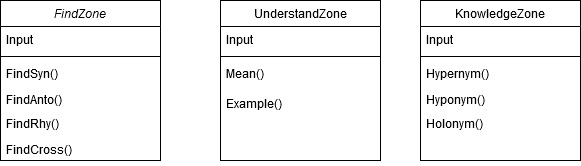
Modules :



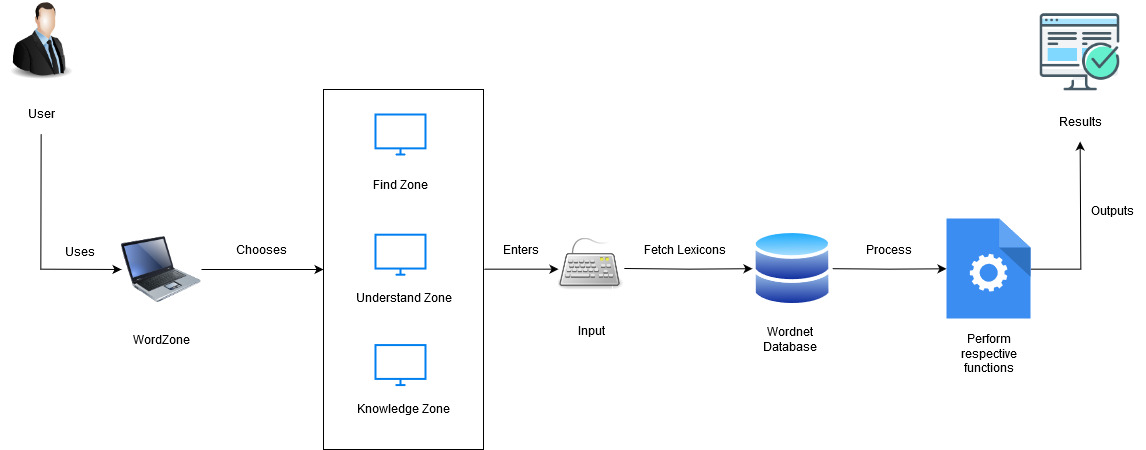
Workflow Diagram:



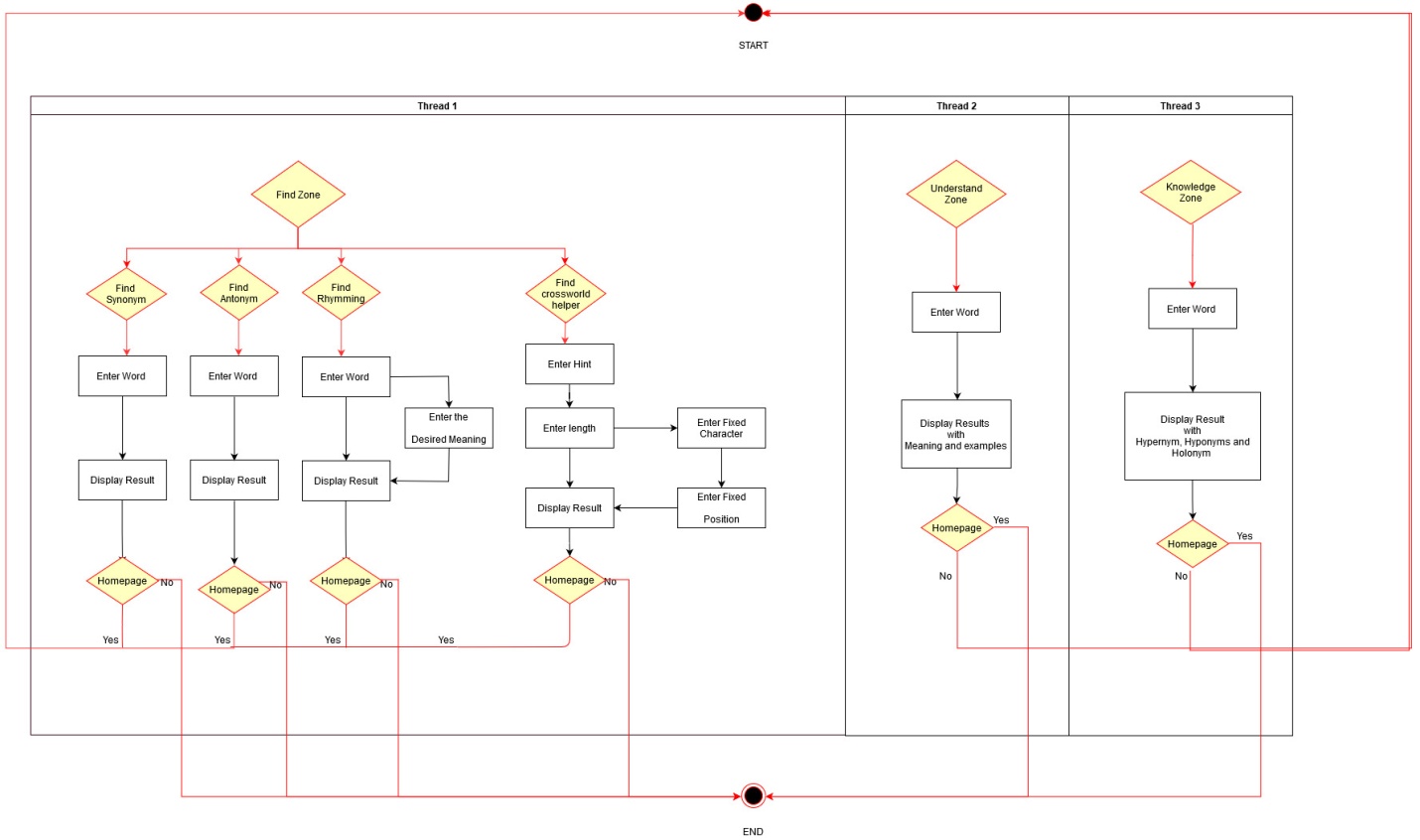
Class Diagram:



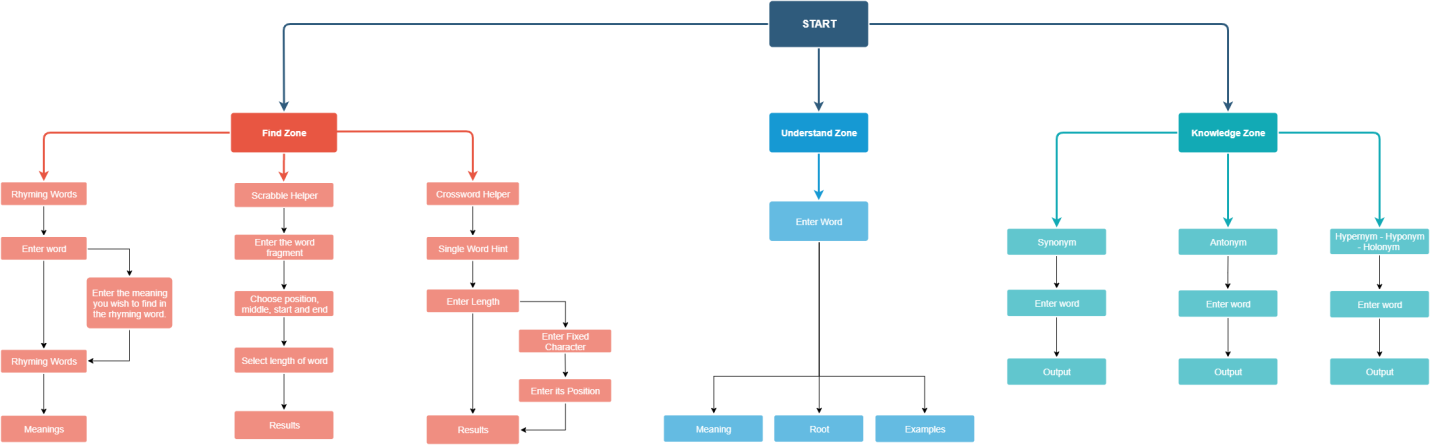
System Architecture:



UML Diagram:



UPDATED DESIGN:



**References:**

[1].Liu Vinci &Curran James(2006),Web Text Corpus for Natural Language Processing.

[2] Loper, Edward & Bird, Steven. (2002). NLTK: the Natural Language Toolkit. CoRR. cs.CL/0205028. 10.3115/1118108.1118117.

[3]Miller, George & Beckwith, R. &Fellbaum, Christiane & Gross, Derek & Miller, Katherine. (1991). Introduction to WordNet: An On-line Lexical Database\*. 3. 10.1093/ijl/3.4.235.

[4]Rajani S, M. Hanumanthappa, 2016, “Techniques of Semantic Analysis for Natural Language Processing – A Detailed Survey”

[5] Dr. M Hanumanthappa, Rashmi S, Jyothi N M, “Impact of Phonetics in Natural Language Processing: A Literature Survey”, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 3, May 2014.

[6]PrimozPodrzaj . A brief demonstration of some Python GUI libraries Proceedings of The 8th International Conference on Informatics and Applications ICIA2019, Japan, 2019

# [7] Hassanin M. Al-Barhamtoshy, Fatimah M. Mujallid,“Building Mobile Dictionary” 2013

# [8]Byrd, Roy &Chodorow, Martin. (2002). Using An On-Line Dictionary To Find Rhyming Words And Pronunciations For Unknown Words. 10.3115/981210.981244.

[9] for wordlist - <https://github.com/dwyl/english-words>