# TTDS Project Plan

December 17, 2021

## 1 Introduction

Our project is to build a **Reversed Dictionary Search Engine**. A similar website is available here:

`https://www.onelook.com/reverse-dictionary.shtml`

Basically, users can type some description/keywords about a single English word as input, and the search engine will return a ranked list (topN) of most related words to the input. For example (from onelook website):

Input: "something heat the food"

Return:

1. oven 2. microwave 3. cookery 4. cooking 5. preparation 6. reheat 7. bake 8. cook 9. roast 10. grill 11. freeze 12. fuel 13. lap 14. action ...

If above basic functionality can be implemented with enough time remaining, possible additional returns can help users to distinguish between topmost two (or more) words.

Addition return:

*"Microwave is just short for "microwave oven" and is one kind of oven. Microwave oven uses microwave radiation to heat food."*

## 2 Motivation

- Help non-English speakers learn more vocabulary.
- Help users solve the problem of *tip-of-the-tongue*[1]

## 3 Team Members

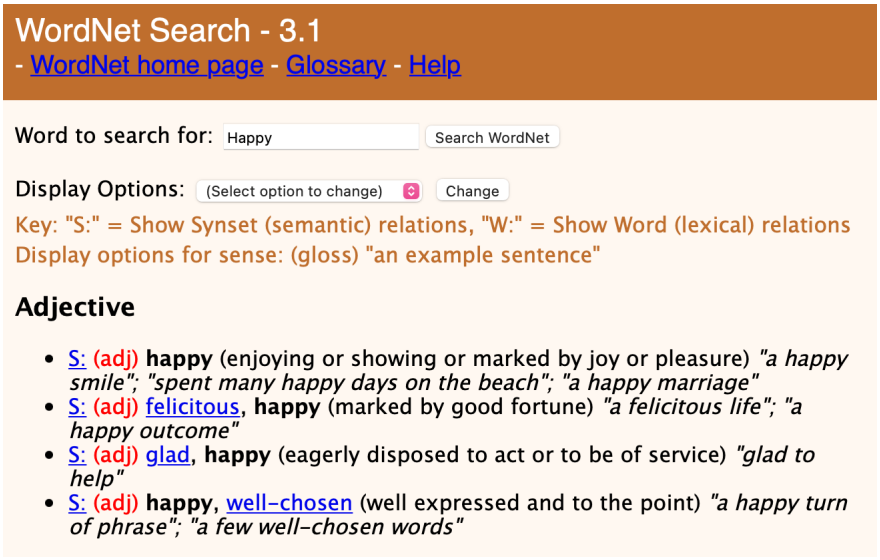
- Youguang Zhou (s2201023)
- Huajian Zhang (s2227123)
- Ruoyi Yu (s2181173)

[1] `https://en.wikipedia.org/wiki/Tip_of_the_tongue`

- Yitian Wang (s2171454)
- Hongyan Deng (s2193724)
- Haochen Qin (s2255142)

# 4 Dataset

Corpus: Wordnet. Each document is all the explanations about a English word like:




We may also use hyponymy relations in Wordnet to deal with semantic relationship between words.

# 5 Baseline Methodology

The simple baseline is just to create a Python dictionary of all words with their explanations (splitted into an array by space). The searching process is to find the matched words in that array.

# 6 Questions

Here are some questions after our first group meeting, could you please kindly give us some feedback? We will be highly appreciated!

- The word explanations in Wordnet are too formal and limited. In order to achieve something like contextualized meaning of a word, do we need more larger corpus?
- Therefore, there will be many modules in our project outside the field of TTDS (e.g. word embedding, POS tagging). Is it suitable or allowed?
- We were wondering the detailed rubric. For example, which one is more important, some interesting, advanced features, or specific indictors, such as effectiveness or F1 score?
- For one possible measurement, can we measure outcome based on pair-wise or list-wist approaches such as defining some loss functions to minimize the nDCG?
- Although our goal is to construct a search engine system, can we use pipelines resembling recommendation system such as using recall&ranking scheme?

# 7 Goal

Finally, we are an ambitious team and our goal is to win this year's best project award!