Alexia Leong, Lily Yan Mr. Brooks

Artificial Intelligence Period 9 Due 11/4/18

Ideas on Word-Connection

1. Is there a way to determine if a pair of words will have a word ladder (if they are not already neighbors)?
2. What about if a pair of words will have a long or short word ladder?

Exploring Idea #1

* We chose to focus on four-letter words only. Doing every single word would take too much time.
* First, we found word ladders for 1000 randomly chosen pairs of four-letter words and recorded their lengths. We also recorded whether or not a word ladder was actually found. We did this procedure five times.
* We looked at the pairs of words without word ladders, and tried to see if there were any patterns among them (i.e. how the words in the pair are related).
* From there, we noticed that some words kept resulting in no word ladders. We recorded those words and tried to find patterns or similarities between them.
* Words we explored:
  + Ache
  + Afar
  + Agar
  + Ajar
  + Ahoy
  + Anal
  + Bevy
  + Ciao
  + Date
  + Ebbs
  + Egad
  + Ends
  + Expo
  + Hymn
  + Imam
  + Judo
  + Luau
  + Meow
  + Nevi
  + Obey
  + Odor
  + Ogre
  + Onto
  + Opal
  + Rhea
  + Semi
* Data:
  + Trial 1
    - CSV: <https://goo.gl/dqGG59>
    - Recurring Words: <https://goo.gl/JLnpJM>
  + Trial 2
    - CSV: <https://goo.gl/3aAetm>
    - Recurring Words: <https://goo.gl/pCQpC9>
  + Trial 3
    - CSV: <https://goo.gl/39tTXC>
    - Recurring Words: <https://goo.gl/ZyDr8H>
* Patterns Observed:
  + Words that start with vowels or have 2-3 vowels tend to not form word ladders.
  + Words with the structure vowel-consonant-vowel-consonant also tend to not form word ladders.
  + There seems to be no noticeable way to tell if two words will form a word ladder.

Exploring Idea #2

* We did the same procedure as above, except this time we used a max heap in our code in order to get long word ladders, and we did 250 random pairs instead of 1000 to cut down on runtime. We also recorded the lengths of the word ladders.
* Data:
  + Trial 1 CSV: <https://goo.gl/StjFvu>
  + Trial 2 CSV: <https://goo.gl/SKE3Cv>
  + Trial 3 CSV: <https://goo.gl/vxAxcf>
* Patterns Observed:
  + The word pairs with the longest paths were:
    - ('cold', 'mans') 1930 pairs
    - ('fuck', 'pans') 1905 pairs
    - ('boon', 'lame') 1901 pairs
    - ('digs', 'less') 1895 pairs
    - ('ever', 'raze') 1894 pairs
    - ('seem', 'dart') 1891 pairs
    - ('ados', 'oars') 1888 pairs
    - ('runt', 'pubs') 1888 pairs
    - ('rags', 'lips') 1888 pairs
    - ('dona', 'jays') `1886 pairs
    - ('coif', 'cars') 1885 pairs
    - ('case', 'hops') 1884 pairs
  + It seems like pairs with only one vowel in each word will have longer word ladders
  + The longest word ladders hover around 1900 words long
  + The words in almost all pairs don’t start with the same letter
  + Most pairs of words don’t have any matching letters in the same position in the words of the pairs
  + Both words in most of the pairs also follow a consonant-vowel-consonant-consonant structure
  + However, there doesn’t seem to be a relationship between two pairs of words that make the word ladder definitively a very long one