Re: Beehive 7-letter words

Received: **March 13, 2016 1:04 PM**

From: **Bill Gilliss bill.gilliss@louisville.edu**

To: **Aaronsdevera aaronsdevera@protonmail.ch**

1 *Attachment* (6.33 MB)

Aaron -

The main word list I have been using has about 173,00 words derived from a longer list developed by the Corpus of Contemporary American English (COCA) and available (or used to be) from www.wordfrequency.info. The attached file is a subset of that list. I can't vouch for *all* the words.

I've attached an Excel file with 38,060 words that each reduce to seven letters after the duplicates are removed. In the DUPLICATES REMOVED column, you can see the seven letters for each. As delivered, the file is sorted on that column, so all the three-point words that use the same seven letters are next to each other. I call words that use the same seven letters a set, and there are 15,568 such sets.

To make these easier to pick out, the THREE-POINT WORDS column identifies such matches, where a 2 indicates that two adjacent words share the same letters, an 8 that eight words do. The most is 87 for AEINRST -- seven of the eight most common letters in English.

I've also included the letter frequency of each letter (based on the frequency chart on Wikipedia), highlighted the lowest one in red (LL), and noted which letter that is. These were originally done with lookups, but I've converted these formulas to values to speed things up. Some columns are hidden.

In creating my own puzzles, I first tried putting the letter with the lowest frequency in the center, but that proved very restrictive for Z, Q, X, J, and K without a C, so when these are in the mix I've been using the letter with the second lowest (2LL) frequency (also included) in the center, and this has made the game much more playable.

As you note, a good puzzle is more than just an algorithm. If this work can advance yours, I am delighted to help.

-Bill Gilliss

On 3/13/2016 11:32 AM, Aaronsdevera wrote:

> Thanks for the message Bill!
> I have all-letter usage further down the development roadmap, but you bring up some good points. If you have code or the wordlist to share, it'd be very helpful to take a look!
>
> Currently my approach has been training the program that generates Beehive to know what a "good" puzzle would be. The public release of the Beehive program, which is the one that publishes to Twitter, has been underperforming hilariously.
>
> Additionally, while Longo certainly employs limits based on letter frequency, letter frequency is certainly only half of the algorithm to generate a "good" puzzle; eg. the March 12 puzzle had an average letter frequency of 7.04 and contained only a handful of solutions while the March 10 puzzle had an average letter frequency of 6.82 and contained many more solutions.
>
> Kind of demonstrates the Longo's cleverness to generate puzzles as a function of letter frequency and letter usability.
>
> -Aaron
>
> -------- Original Message --------
> Subject: Beehive 7-letter words
> Local Time: March 13, 2016 12:58 AM
> UTC Time: March 13, 2016 5:58 AM
> From: bill.gilliss@louisville.edu
> To: aaronsdevera@protonmail.ch
>
> Hey, Aaron -

I greatly enjoy Frank Longo's Beehive in the Times Magazine, and have just discovered your Daily Beehive Twitter page. What a pleasure! My great delight was dampened only by discovering that there was not necessarily a seven-letter solution to each puzzle, the three-pointer I strive to find each week in Frank's puzzle.

I've been working on anagram-related programming recreationally over the years, for an app that never came into being, and could readily provide thousands of sets of letters for your Daily Beehive puzzle that each had at least one seven-letter solution.

My list of words with exactly seven unique letters has about 15,600 distinct sets. Many such sets have two or three such solutions that use all the letters at least once, some have a dozen, one has 87! The set INORSTU, for instance, has three: NITROUS, NUTRITIOUS, NOTORIOUS

I have paid more attention to word-frequency analysis than to letter-frequency analysis. Perhaps your Python code could pick through my sets of letters to find those that match the frequency that Frank uses.

I have not been concerned at all with finding the many shorter words that can be composed with each set (which is to say, with solving Beehive computationally), but there is a lot of software that does that -- perhaps you use some already.

Anyway, if a quarter of the seven-letter sets have unique solutions that are just too obscure to inflict upon the public, like PROTYLS or EMBRYON, and another quarter will have such a high letter-frequency score that there would be too *many* solutions, that would still leave 8,000 usable sets. Call it 20 years' worth.

Wouldn't it be fun for people to find the following three-pointers in the same set of letters?
ACENTRIC
ANCIENTER
ANTICANCER
CENTENARIAN
CERATIN
CERTAIN

CIRCINATE
CRANIATE
CREATINE
CREATININE
INCARCERATE
INCARNATE
INCINERATE
INTERACT
INTERACTANT
INTRICATE
NECTARINE
REINCARNATE
TACRINE

I've got lots more. Interested?

-Bill Gilliss
bill.gilliss@louisville.edu