FiveThirtyEight's January 10, 2020 Riddler

Emma Knight

January 15, 2020

This is my solution to the riddler from January 10, 2020.

Before even stating the question, I first need to define a couple of things:

Definition 1.
- *The alphanumeric value of a character (technically, an unaccented character in the Latin alphabet) is the position of said character in the alphabet. For example, the alphanumeric value of the character "f" is 6.*

- *The alphanumeric value of a set of words is the sum of the values of the characters that comprise the words. For example, the alphanumeric value of "red hot" is $18 + 5 + 4 + 8 + 15 + 20 = 70$.*

- *The alphanumeric value of a nonnegative integer is value of the English word for that number (with no "and"s). For example, the alphanumeric value of 123, which is just the alphanumeric value of "one hundred twenty three," is 271.*

With that in mind, we can now ask the question:

Question 2. *What is the largest integer n such that n is less than the alphanumeric value of n?*

A quick heuristic argument shows that n has $\ln(n)$ digits, each of which needs $O(1)$ characters to give the value of the digit, and $O(\ln(\ln(n)))$ characters to give the position in the number, so n needs $O(\ln(n)\ln(\ln(n)))$ characters to write out. Since the value of a character is also $O(1)$, the value of n is $O(\ln(n)\ln(\ln(n)))$. Now, $\ln(n)\ln(\ln(n))$ grows much more slowly than n, so this number is well defined. Precise bounds, however, are useful in knowing that you are done.

Lemma 3. *For $n \geq 1000$, the value of n is at most $(26\ln(n))(34 + 10\ln(\ln(n)))$.*

Proof. Write $k = 1 + \lfloor \log_{1000}(n) \rfloor$. Then n, in words, is given by k three-digit numbers followed by a delimiter (i.e. 12345678 is "twelve million, four hundred fifty six thousand, seven hundred eighty nine," with delimiters "million," "thousand," and "" respectively). The length of a three-digit number as a word is at most 24: "seven" is the longest digit and "seventy" is the longest tens digit, so the worst case is "seven hundred seventy seven" which has 24 characters. The delimiter (using Conway-Guy notation for large numbers) has size at most $10(1 + \log_{10}(k))$. Thus, there are at most $k(34 + 10\log_{10}(k))$ digits, and so the alphanumeric value of n is at most $26k(34 + 10\log_{10}(k))$. Now, using $k < \ln(n)$ and $\log_{10}(k) < \ln(k)$, one gets the claimed inequality. $\square$

I would like to state that this approximation is atrocious: The largest alphanumeric value of any number less than 1000000 is the alphanumeric value of 777777, which is 730. Conversely, one has that $f(777777) = 23713.051\ldots$. Anyways, let $f(x) = 26\ln(x)(34 + 10\ln(\ln(x)))$. To see that the number asked for in the question is well-defined, we need one more lemma:

Lemma 4. *One has that $f(x) < x$ for $x \geq 1000000$.*

Proof. A quick computation gives $f(1000000) = 21644.84187\ldots$, which is well less than 1000000. One can also compute $f'(x) = \frac{26(44 + 10\ln(\ln(x)))}{x}$. Plugging in $f'(1000000)$, one gets $.0252\ldots$, a number which is well less than 1. Additionally, its visible that the numerator is growing slower than the denominator, so for $x \geq 1000000$ one has that $f'(x) < 1$ and so, since $f(1000000) < 1000000$, one sees that $f(x) < x$ for $x \geq 1000000$. $\square$

Thus, any numbers which are less than their alphanumeric value are at most 999999. This is an aggressively finite problem, and at this point one can just wrote code to find all of the examples of such numbers. Running the code, one sees that there are 254 such numbers, the largest of which is 279 with an alphanumeric value of 284.

I believe that its possible to sharpen the estimation here and bound things at 1000, but since iterating such a simple operation over one million numbers is well within what modern computers are capable of I found no reason to do so.

It is also simple to change this question to other sorts of values associated to words. Another simple function that many are familiar with is the (English) Scrabble value, where each letter is just assigned it's score as a Scrabble tile. Unsurprisingly, the largest number that is smaller than it's Scrabble value is a lot smaller than the largest number that is smaller than it's alphanumeric value. I found it a little surprising that it was only 8, with a Scrabble value of 9. This was smaller than I expected (although it makes sense in retrospect).