Homework 6: Exercise 8.14

Maryam Bahrani (mbahrani)

David Luo

3/5

Index the characters using an arbitrary enumeration, denoting them as atoms $\mathcal{Z}_i$.

Denote the given pattern by P, and let $P(i)$ be the number corresponding the ith character of the pattern. Note that the given pattern has $l = 44$ characters. Let $\mathcal{K}$ be the class of strings that do not include the given pattern.

For any string $k \in \mathcal{K}$, we consider the length of the longest common prefix of k and the pattern, giving l possibilities (length of longest common suffix could be $0, 1, \cdots, l-1$:

If the length of longest common prefix is i, the next character is either empty, or any of the 31 characters other than $P(i)$ (to ensure not having a longer common prefix) followed by an element of P. Therefore, k is symbolically;

$$\mathcal{Z}_{P(1)} \times \mathcal{Z}_{P(1)} \times \cdots \mathcal{Z}_{P(i)} \times (\epsilon + \sum_{j \neq i} \mathcal{Z}_j \times \mathcal{K}),$$

This includes strings that contain the pattern. For example if the prefix is THE and the next letter is T then an element of P is HE QUICK BROWN FOX etc

which translates to the following terms in the generating function

$$z^i(1 + 31zK(z)).$$

The generating function for $\mathcal{K}$ is the sum of the above term over all values of i: -2

$$K(z) = \sum_{i=0}^{l-1} z^i(1 + 31zK(z)) = (1 + z + \cdots + z^{l-1})(1 + 31zK(z))$$

$$K(z) = \frac{1 + z + \cdots + z^{l-1}}{1 - 31(z + z^2 + \cdots + z^l)} \cdots \frac{1 - z^l}{1 - 32z + 31z^{l+1}}$$

As shown in lecture, the expected end position of the first occurrence of the pattern is the generating function evaluated at 1/(size of alphabet). This is the case because

$$\begin{aligned} K(1/32) &= \sum_{N \geqslant 0} \frac{(\text{ of strings with no runs of 32 zeros})}{32^N} \\ &= \sum_{N \geqslant 0} \text{Probability that first } N \text{ bits of a random string have no occurrence of pattern} \\ &= \sum_{N \geqslant 0} \text{Expected position of end of first occurrence of pattern} \end{aligned}$$

Therefore, the expected length of a strings required to get an occurrance of the pattern is roughly

$$K(1/32) = 1.74 \cdot 10^{66}.$$