Combinatorics on Words

Woong Choi
woong250606@gmail.com

Euler Circle

July 23, 2023

Table of contents

Definition of a Word

A word is a permutation of a series of alphabets. Because some words have very intriguing properties, I will start by defining a few definitions that are critical to understand later concepts.

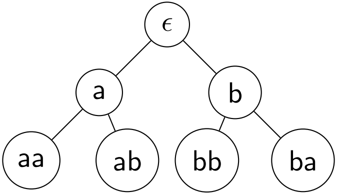
Definition

Semigroup is a set under associative binary operation. Subsemigroup is a subset closed under the binary operation.

Monoid and Tree Diagram

Definition

A set M is called monoid if M contains a neutral element ϵ such that mϵ=ϵm=m for m $\in$ M. Submonoid is a subset of M such that it is closed under the operation and contains ϵ.



Factor of a Word

Definition

A word w is called a factor of u if $\exists$ words a,b such that $u = awb$. w is a *proper factor* if $ab \neq \epsilon$. Furthermore, the set of all factors of w is denoted as $F(x)$.

"is","nikashisfast" are both factors of "nikashisfast"

Fine and Wilf's Theorem

Theorem

For a word w with two periods p and q, if $|w| \geq p + q - gcd(p, q)$, then w has the period $gcd(p, q)$.

Stronger Version

Theorem

Let w be a word with periods p_1, p_2, p_3 satisfying $p_1 \leq p_2 \leq p_3$. If $|w| \geq f(p_1, p_2, p_3)$, w also has the period $gcd(p_1, p_2, p_3)$.

Fine and Wilf's Theorem Cont'

Definition

If x is a prefix of y, x is a substring of y taken from index 0 to i with $i \leq |y| - 1$. Prefix of length j of y is denoted as $\text{pref}_j(y)$.

Definition

If x is a suffix of y, x is a substring of y taken from index i to $|y| - 1$ with $i \geq 0$. Suffix of length j of y is denoted as $\text{suff}_j(y)$.

Definition

$h(p) = |p^{m(p)}|$

where $m_p = min\{k \mid p_1^k = 0\}$.

Fine and Wilf's Theorem Cont'

Definition

$f(x, y, z) = 1/2(x + y + z - 2gcd(x, y, z) + h(x, y, z))$ where x,y,z are nonnegative integers.

Fine and Wilf's Theorem Cont'

Lemma

Let $p, q \in \mathbb{Z}^+$ *satisfying* $p < q$. *Let* w *be a word such that* $|w| = n$ *and has two periods* p *and* q. *Then, the prefix and suffix of* w *of length n-p has period q-p.*

Fine and Wilf's Theorem Cont'

Lemma

Let w be a word with $|w| = n$ having periods p, $q \in \mathbb{Z}^+$. Let $j = \text{pref}_p(w)w$. j has periods p and $p+q$.

Fine and Wilf's Theorem Cont'

Restatement of the main theorem:

Theorem

Let w be a word with periods p_1, p_2, p_3 satisfying $p_1 \leq p_2 \leq p_3$. If $|w| \geq f(p_1, p_2, p_3)$, w also has the period $gcd(p_1, p_2, p_3)$.

Proof.

Let $w = uv, |u| = p_1, |v| = |w| - p_1$. By lemma 3.12, v has periods $p_1, p_2 - p_1, p_3 - p_1$. $|v| = |w| - p_1 \geq f(p_1, p_2, p_3) - p_1 = 1/2(p_1 + p_2 - p_1 + p_3 - p_1 - 2gcd(p_1, p_2, p_3) + h(p_1, p_2, p_3) = f(p_1, p_2 - p_1, p_3 - p_1)$ because $gcd(p_1, p_2, p_3) = gcd(p_1, p_2 - p_1, p_3 - p_1)$ and same holds for $h(p_1, p_2, p_3)$ because of Euclidean algorithm. ∎

Sturmian Word

Definition

A word w is called a Sturmian word if $|w|$ is infinite, is composed of only two distinct alphabets, and contains $n+1$ factors for each $n \geq 0$.

Example of Sturmian Word

0100101001001. . .

Special Thanks

I want to thank Simon, Carson, and everyone that was supportive when I asked for help in Euler circle. Everyone's enthusiasm for math in this class helped me to understand all the concepts very well. I will show appreciation by ending the presentation with a duck.