Pattern Avoidance in Words

Anshul Mantri

Euler Circle

July 14, 2024

Basic Definitions

Basic Definitions

Definition

An *alphabet* is a finite set of symbols, and is denoted by A. The elements of A are called *letters*.

Basic Definitions

Definition

An *alphabet* is a finite set of symbols, and is denoted by A. The elements of A are called *letters*.

Definition

A *word* is a sequence of letters from an alphabet A.

Basic Definitions

Definition

An *alphabet* is a finite set of symbols, and is denoted by A. The elements of A are called *letters*.

Definition

A *word* is a sequence of letters from an alphabet A.

Definition

A *factor* or *subword* of a word w is a contiguous subsequence of w of any length.

More Definitions

More Definitions

Definition

The *free monoid* of an alpabet A, denoted as A^*, is the set of all words of any length that can be made with the letters of A.

This set is a monoid because it is closed under the associative operation of concatenation, and it has the empty word ε as the identity element.

More Definitions

Definition

The *free monoid* of an alpabet A, denoted as A^*, is the set of all words of any length that can be made with the letters of A.

This set is a monoid because it is closed under the associative operation of concatenation, and it has the empty word ε as the identity element.

Definition

A *morphism* h from A^* and B^* is a function that maps each element of A^* to an element of B^*, such that for all $m, n \in A^*$, we have $h(m)h(n) = h(mn)$.

Patterns

Patterns

Definition

A *pattern* is a word where the letters are variables.

Each variable represents a nonempty word in A^*.

Patterns

Definition

A *pattern* is a word where the letters are variables.

Each variable represents a nonempty word in A^*.

Definition

A word w *encounters* a pattern p if it contains a factor that can be made by substituting each variable in p with a word in A^*. On the other hand, if a w does not encounter a pattern, w *avoids* the pattern.

Patterns

Definition

A *pattern* is a word where the letters are variables.

Each variable represents a nonempty word in A^*.

Definition

A word w *encounters* a pattern p if it contains a factor that can be made by substituting each variable in p with a word in A^*. On the other hand, if a w does not encounter a pattern, w *avoids* the pattern.

Definition

If there exists an infinite word on an alphabet with k letters that avoids a pattern p, then p is *k-avoidable*. Otherwise, p is *k-unavoidable*.

Thue-Morse Morphism

Let $A = \{a, b\}$. Define the morphism $\mu : A^* \to A^*$, where

$$\mu(a) = ab$$
$$\mu(b) = ba$$

Thue-Morse Morphism

Let $A = \{a, b\}$. Define the morphism $\mu : A^* \to A^*$, where

$$\mu(a) = ab$$
$$\mu(b) = ba$$

We can generate the infinite word t by repeatedly applying μ to the starting letter a.

$$\begin{aligned} t_0 &= a \\ t_1 = \mu(t_0) &= ab \\ t_2 = \mu(t_1) &= abba \\ t_3 = \mu(t_2) &= abbabaab \\ &\vdots \\ t = \mu(t) &= abbabaabbaababba\ldots \end{aligned}$$

Thue-Morse Word

The infinite word t is called the Thue-Morse word.

Proposition

The Thue-Morse word avoids the following patterns:

Thue-Morse Word

The infinite word t is called the Thue-Morse word.

Proposition

The Thue-Morse word avoids the following patterns:

1. xxx

Thue-Morse Word

The infinite word t is called the Thue-Morse word.

Proposition

The Thue-Morse word avoids the following patterns:

1. xxx
2. $xyxyx$

Thue-Morse Word

The infinite word t is called the Thue-Morse word.

Proposition

The Thue-Morse word avoids the following patterns:

1. xxx
2. $xyxyx$

Proof (outline).

FTSOC, assume t encounters the pattern, so there is a factor of t with minimal length that encounters the pattern. Then use the fact that μ maps t to itself to show that there is a factor following the pattern that is half the length, which is a contradiction. ∎

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$
$Z_0 = \varepsilon$

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$
$Z_0 = \varepsilon$
$Z_1 = a$

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$
$Z_0 = \varepsilon$
$Z_1 = a$
$Z_2 = aba$

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$
$Z_0 = \varepsilon$
$Z_1 = a$
$Z_2 = aba$
$Z_3 = abacaba$

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$
$Z_0 = \varepsilon$
$Z_1 = a$
$Z_2 = aba$
$Z_3 = abacaba$
$Z_4 = abacabadabacaba$

Sesquipowers

Definition

The *sesquipowers*, also called the *Zimin words*, are defined as follows. Let $Z_0 = \varepsilon$, and for every $\sigma \in |\Sigma|$, let σ_n be a new symbol in an alphabet Σ. Then, $Z_{n+1} = Z_n \sigma_n Z_n$.

Example

$\Sigma = \{a, b, c, d\}$
$Z_0 = \varepsilon$
$Z_1 = a$
$Z_2 = aba$
$Z_3 = abacaba$
$Z_4 = abacabadabacaba$

We are interested in the Zimin patterns, which are sesquipowers made of variables instead of letters. Since there can be arbitrarily many variables in a pattern, there are infinitely many Zimin patterns.

Zimin Patterns are Unavoidable

Proposition

Every Zimin pattern is unavoidable on all alphabets.

Zimin Patterns are Unavoidable

Proposition

Every Zimin pattern is unavoidable on all alphabets.

The proof is by induction. The base case is trivial.

Lemma

Let p be a pattern that is unavoidable on an alphabet A. If x is a variable that does not appear in p, then the pattern pxp is also unavoidable on A.

Zimin Patterns are Unavoidable

Proposition

Every Zimin pattern is unavoidable on all alphabets.

The proof is by induction. The base case is trivial.

Lemma

Let p be a pattern that is unavoidable on an alphabet A. If x is a variable that does not appear in p, then the pattern pxp is also unavoidable on A.

Proof.

There must exist some finite length l such that all words of length l encounter p. If we have sufficiently many blocks of l letters, separated by one letter, then some word of length l is repeated. Then x is the word between the two occurence of p, and we have the pattern pxp. ■

Zimin's Theorem

Theorem (Zimin, 1984)

A pattern is unavoidable on all alphabets if and only if it is a factor of a Zimin pattern.

To prove this, Zimin showed that a pattern is unavoidable if and only if it is reducible using the Zimin algorithm.

The End

Thanks for listening!

If you found this interesting, please check out my paper for more.