A glimpse of combinatorics on words

Benny George K

Tata Institute of Fundamental Research,Mumbai
India

Indo Russian Workshop 2008
Ural State Univeristy

Outline

1. Introduction

2. Theorem Gallery
   - Fine and Wilf's Theorem
   - Pattern Avoidance Theorems

3. Equation over words

Outline

1. Introduction
2. Theorem Gallery
   - Fine and Wilf's Theorem
   - Pattern Avoidance Theorems
3. Equation over words

Outline

1. Introduction
2. Theorem Gallery
   - Fine and Wilf's Theorem
   - Pattern Avoidance Theorems
3. Equation over words

Outline

1. Introduction
2. Theorem Gallery
   - Fine and Wilf's Theorem
   - Pattern Avoidance Theorems
3. Equation over words

Basic Terminology

- A is a finite alphabet. A^+ (resp. A^*)is the free semigroup (resp. monoid) generated by A
- Elements of A^* are called *words*. Subsets of A^* are called *languages*.
- Lower cases letters $x, y, z, \ldots$ denotes words and upper case letters $X, Y, Z \ldots$ denotes languages. Lower case letters $a, b, c, \ldots$ are used for constants.
- 1 denotes the empty word.
- $|w|$ is length of word w.
- For two words u and v, u is a *prefix*(*suffix*, *factor*) of v if there exists x, y such that $v = ux(v = xu, v = xuy)$.

Basic Terminology

- A is a finite alphabet. A^+ (resp. A^*)is the free semigroup (resp. monoid) generated by A
- Elements of A^* are called *words*. Subsets of A^* are called *languages*.
- Lower cases letters $x, y, z, \ldots$ denotes words and upper case letters $X, Y, Z \ldots$ denotes languages. Lower case letters $a, b, c, \ldots$ are used for constants.
- 1 denotes the empty word.
- $|w|$ is length of word w.
- For two words u and v, u is a *prefix*(*suffix*, *factor*) of v if there exists x, y such that $v = ux (v = xu, v = xuy)$.

Basic Terminology

- A is a finite alphabet. A^+ (resp. A^*)is the free semigroup (resp. monoid) generated by A
- Elements of A^* are called *words*. Subsets of A^* are called *languages*.
- Lower cases letters $x, y, z, \ldots$ denotes words and upper case letters $X, Y, Z \ldots$ denotes languages. Lower case letters $a, b, c, \ldots$ are used for constants.
- 1 denotes the empty word.
- $|w|$ is length of word w.
- For two words u and v, u is a *prefix*(*suffix*, *factor*) of v if there exists x, y such that $v = ux (v = xu, v = xuy)$.

Basic Terminology

- A is a finite alphabet. A^+ (resp. A^*)is the free semigroup (resp. monoid) generated by A
- Elements of A^* are called *words*. Subsets of A^* are called *languages*.
- Lower cases letters $x, y, z, \ldots$ denotes words and upper case letters $X, Y, Z \ldots$ denotes languages. Lower case letters $a, b, c, \ldots$ are used for constants.
- 1 denotes the empty word.
- $|w|$ is length of word w.
- For two words u and v, u is a *prefix*(*suffix*, *factor*) of v if there exists x, y such that $v = ux(v = xu, v = xuy)$.

Basic Terminology

- A is a finite alphabet. A^+ (resp. A^*)is the free semigroup (resp. monoid) generated by A
- Elements of A^* are called *words*. Subsets of A^* are called *languages*.
- Lower cases letters $x, y, z, \ldots$ denotes words and upper case letters $X, Y, Z \ldots$ denotes languages. Lower case letters $a, b, c, \ldots$ are used for constants.
- 1 denotes the empty word.
- $|w|$ is length of word w.
- For two words u and v, u is a *prefix*(*suffix*, *factor*) of v if there exists x, y such that $v = ux(v = xu, v = xuy)$.

Basic Terminology

- A is a finite alphabet. A^+ (resp. A^*)is the free semigroup (resp. monoid) generated by A
- Elements of A^* are called *words*. Subsets of A^* are called *languages*.
- Lower cases letters $x, y, z, \ldots$ denotes words and upper case letters $X, Y, Z \ldots$ denotes languages. Lower case letters $a, b, c, \ldots$ are used for constants.
- 1 denotes the empty word.
- $|w|$ is length of word w.
- For two words u and v, u is a *prefix*(*suffix*, *factor*) of v if there exists x, y such that $v = ux(v = xu, v = xuy)$.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^{\infty}(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order* k
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \dots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^\infty(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order k*
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \ldots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^\infty(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order* k
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \ldots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^\infty(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order* k
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \ldots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^{\infty}(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order* k
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \dots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^\infty(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order* k
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \dots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Basic Terminology

- h is a *morphism* from A^* to B^* if $h(uv) = h(u)h(v)$
- $\phi(a) = ab, \phi(b) = a$ This is called *fibonacci* morphism. Let $f = \phi^\infty(a)$
- $f_0 = a, f_1 = ab$ and $f_{n+1} = f_n f_{n-1}$.
- If $w = xu^k y$, we say w has a *repetition of order k*
- If a words doesn't contain a repetition of order k its is called $k - free$.
- A word is called *primitive* if it is not of the form w^k for $k > 1$.
- An integer p is a *period* of $w = a_1 \ldots a_n$ if $a_i = a_{p+i}$. The smallest period is often called *the period*.

Outline

Fine and Wilf's Theorem

Let w be a word of length n. If w has two period p and q and $n \geq p + q - gcd(p, q)$, then $gcd(p, q)$ is a period of w.

- Bound is tight. Consider $w = abaab\ abaab\ a$. It is a word of length 11 and is not power of a single letter.
- Such words are called *Sturmian* words.

Fine and Wilf's Theorem

Let w be a word of length n. If w has two period p and q and $n \geq p + q - gcd(p, q)$, then $gcd(p, q)$ is a period of w.

- Bound is tight. Consider $w = abaab\ abaab\ a$. It is a word of length 11 and is not power of a single letter.
- Such words are called *Sturmian* words.

Fine and Wilf's Theorem

Let w be a word of length n. If w has two period p and q and $n \geq p + q - gcd(p, q)$, then $gcd(p, q)$ is a period of w.

- Bound is tight. Consider $w = abaab\ abaab\ a$. It is a word of length 11 and is not power of a single letter.
- Such words are called *Sturmian* words.

Square free and Cube free words

Find as long as possible (infinite?) word over an n letter alphabet that it is $k - free$.

Square free and Cube free words

Find as long as possible (infinite?) word over an n letter alphabet that it is $k - free$.

Thue 1906

There exists an infinite $2^{+} - free$ word over binary alphabet.

Square free and Cube free words

Find as long as possible (infinite?) word over an n letter alphabet that it is $k - \textit{free}$.

Thue 1906

There exists an infinite $2^+ - \textit{free}$ word over binary alphabet.

Thue 1906

There exists an infinite $2 - \textit{free}$ word over three alphabet.

Outline

1. Introduction
2. Theorem Gallery
   - Fine and Wilf's Theorem
   - Pattern Avoidance Theorems
3. Equation over words

An Example

- Consider the following equation
- All its solution are of the form $x = a(ba)^n, y = (ba)^n b$
- We try to solve one such equation,but over languages

An Example

- Consider the following equation

$$bxyx = yxya$$

- All its solution are of the form $x = a(ba)^n, y = (ba)^n b$
- We try to solve one such equation,but over languages

An Example

- Consider the following equation

$$bxyx = yxya$$

- All its solution are of the form $x = a(ba)^n, y = (ba)^n b$
- We try to solve one such equation,but over languages

An Example

- Consider the following equation

$$bxyx = yxya$$

- All its solution are of the form $x = a(ba)^n, y = (ba)^n b$
- We try to solve one such equation,but over languages

Problem Statement

- Conjugacy equation is $XZ = ZY$
- Input: X and Y
- Output: Yes iff there exists a Z.
- We consider the case when $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$

Problem Statement

- Conjugacy equation is $XZ = ZY$
- Input: X and Y
- Output: Yes iff there exists a Z.
- We consider the case when $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$

Problem Statement

- Conjugacy equation is $XZ = ZY$
- Input: X and Y
- Output: Yes iff there exists a Z.
- We consider the case when $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$

Problem Statement

- Conjugacy equation is $XZ = ZY$
- Input: X and Y
- Output: Yes iff there exists a Z.
- We consider the case when $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$

Folklore Lemmas

Lemma 1

If p is a primitive word and $p^2 = xpy$, then either $x = 1$ or $y = 1$.In other words, primitive words cannot occur inside their powers in a nontrivial manner.

Folklore Lemmas

Lemma 1

If p is a primitive word and $p^2 = xpy$, then either $x = 1$ or $y = 1$.In other words, primitive words cannot occur inside their powers in a nontrivial manner.

Lemma 2

If $xz = zy$, then $x = (uv)^r, y = (vu)^r$ and $z = (uv)^{r'}u$ and uv is primitive.

Folklore Lemmas

Lemma 1

If p is a primitive word and $p^2 = xpy$, then either $x = 1$ or $y = 1$.In other words, primitive words cannot occur inside their powers in a nontrivial manner.

Lemma 2

If $xz = zy$, then $x = (uv)^r, y = (vu)^r$ and $z = (uv)^{r'}u$ and uv is primitive.

Lemma 3

If $xy = yx$, then $x = (w)^{r_1}, y = (w)^{r_2}$

Folklore Lemmas

Lemma 1

If p is a primitive word and $p^2 = xpy$, then either $x = 1$ or $y = 1$.In other words, primitive words cannot occur inside their powers in a nontrivial manner.

Lemma 2

If $xz = zy$, then $x = (uv)^r, y = (vu)^r$ and $z = (uv)^{r'}u$ and uv is primitive.

Lemma 3

If $xy = yx$, then $x = (w)^{r_1}, y = (w)^{r_2}$

Lemma 4

If $XZ = ZY$ and X commutes then so does Y.

THANK YOU