Vermeidbare Muster in unendlichen Wörtern

Heiko Stamer

Universität Kassel, Fachbereich Mathematik/Informatik
Heinrich-Plett-Straße 40, D-34132 Kassel

`stamer@theory.informatik.uni-kassel.de`

`76F7 3011 329D 27DB 8D7C 3F97 4F58 4EB8 FB2B E14F`

Schülerprojektwoche 2005

U N I K A S S E L
V E R S I T Ä T

UNI KASSEL
VERSITÄT

Informell: Folgen von Zeichen aus einem gegebenen Zeichenvorrat.

- Worte und Sätze einer natürlichen Sprache

  Zeichenvorrat: a, . . . , z, A, . . . , Z, ␣, ,, ., . . . , ?, !

  Beispiele: `Hallo␣Welt!`, `X` und `WtRfqkZj?`

- Programme einer Programmiersprache

  Zeichenvorrat: von der Eingabekodierung und Programmiersprache abhängig

  Beispiel: (gibt eigenen Quelltext aus)

```
#!/usr/bin/env␣python
def␣p(P):
␣␣return␣reduce((lambda␣P,p:P+"\n"+p),P)
P=['#!/usr/bin/env␣python',␣'def␣p(P):',␣'␣␣return␣reduce((lambda␣P,p
    :P+"\\n"+p),P)',␣'P=',␣'P[3]="P="+`P`',␣'print␣p(P)']
P[3]="P="+`P`
print␣p(P)
```

- Arithmetische Ausdrücke

  Zeichenvorrat: `0`, `1`, `2`, `3`, `4`, `5`, `6`, `7`, `8`, `9`, `.`, `+`, `-`, `*`, `/`, `=`, `(`, `)`

  Beispiele: `7*3=21`, `1=-2` und `0.815++)23=42*`

- . . .

Die *Kombinatorik* untersucht die Struktur von Wörtern (Syntax), ohne ihre eigentliche Bedeutung (Semantik) zu berücksichtigen.

Kombinatorik auf Wörtern (Combinatorics on Words):

- Teilgebiet der diskreten Mathematik
- Hauptsächlich motiviert aus der Informatik, aber auch diverse Querverbindungen in andere Bereiche:
  - Mathematik: Zahlentheorie, dynamische Systeme, Logik
  - Algebra: freie Gruppen, Matrizen, Darstellungen, Burnside Probleme
  - Biologie: Verarbeitung und Suche in DNA-Sequenzen, Lindenmayer-Systeme (Theorie zum Pflanzenwachstum)
  - Informatik: Algorithmik, Datenkompression, Automaten- und Berechenbarkeitstheorie, Mustererkennung, Compilerbau
- Viele kombinatorische Probleme sind einfach zu beschreiben, haben aber komplizierte Lösungen oder sind noch offen.

Definition

Ein (endliches) Alphabet ist eine (endliche) Menge von Symbolen.

Beispiel (Alphabet für arithmetische Ausdrücke)

$\Sigma = \{\texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{.}, \texttt{+}, \texttt{-}, \texttt{*}, \texttt{/}, \texttt{=}, \texttt{(}, \texttt{)}\}$

Definition

Sei Σ ein endliches Alphabet und n eine nicht-negative ganze Zahl.
Ein Wort der Länge n über Σ ist eine Abbildung $w : [1, n] \to \Sigma$.

Schreibweisen

$(w(1), w(2), \ldots, w(n))$, $(w_1, w_2, \ldots, w_n)$, oder $w_1 w_2 \cdots w_n$

Definition

Für eine nicht-negative ganze Zahl $n \in \mathbb{N}$ ist $\Sigma^n = \{w \mid w : [1, n] \to \Sigma\}$ die Menge der Worte der Länge n über Σ.

- $\Sigma^0 = \{w \mid w : \emptyset \to \Sigma\}$ enthält nur das leere Wort ϵ.
- $\Sigma^* = \bigcup_{n \geq 0} \Sigma^n$ ist die Menge aller Worte über Σ.
- $\Sigma^+ = \bigcup_{n \geq 1} \Sigma^n$ ist die Menge der nicht-leeren Worte über Σ.

Beispiel (Wortmengen)

- $\Sigma^0 = \{\epsilon\}$, $\Sigma^1 = \Sigma = \{\texttt{0}, \ldots, \texttt{)}\}$, $\Sigma^2 = \{\texttt{00}, \texttt{01}, \ldots, \texttt{)(}, \texttt{))}\}$, usw.
- $\Sigma^* = \{\epsilon, \texttt{0}, \ldots, \texttt{)}, \texttt{00}, \ldots, \texttt{))}, \texttt{000}, \ldots\}$, $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$

Definition

Eine (endliche) Teilmenge $L \subseteq \Sigma^*$ heißt (endliche) Sprache über Σ.

Definition

Die Länge $|w| : \Sigma^* \to \mathbb{N}$ eines Wortes $w \in \Sigma^*$ ist durch $|\epsilon| = 0$ oder $|(w_1, w_2, \ldots, w_n)| = n$ gegeben.

Definition

Für ein Symbol $z \in \Sigma$ und ein Wort $w \in \Sigma^*$ gibt $|w|_z : \Sigma^* \to \mathbb{N}$ die Anzahl der Vorkommen von z in w an.

Definition

Die Konkatenation $\cdot : \Sigma^* \times \Sigma^* \to \Sigma^*$ (oder auch Verkettung) zweier Wörter $u, v \in \Sigma^*$ entspricht dem „Hintereinanderschreiben“, d. h. $(u_1, u_2, \ldots, u_n) \cdot (v_1, v_2, \ldots, v_m) = (u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_m)$.

Beispiel (u = 7*3=21 und v = /(8))

$|u| = 6$, $|u|_1 = 1$, $|v| = 4$, $|v|_1 = 0$, $u \cdot v$ = 7*3=21/(8), $|u \cdot v| = 10$

Die Konkatenation ist eine binäre Operation auf Σ^*, denn für alle Wörter $u \in \Sigma^n$ und $v \in \Sigma^m$ gilt nach Definition $(u \cdot v) \in \Sigma^{n+m}$.

Lemma (wichtige Eigenschaften der Verkettung)

1. *Das leere Wort ϵ ist sowohl links- als auch rechts-neutral, d. h. $\epsilon \cdot u = u = u \cdot \epsilon$.*
2. *Assoziativität, d. h. $(u \cdot v) \cdot w = u \cdot (v \cdot w)$ für alle $u, v, w \in \Sigma^*$.*
3. *Kommutativität herrscht nur genau dann, wenn Σ unär ist.*
4. *Für beliebige $u, v, w \in \Sigma^*$ gelten folgende Kürzungsregeln:*
   - *$u \cdot v = u \cdot w$ impliziert $v = w$,*
   - *$u \cdot w = v \cdot w$ impliziert $u = v$.*
5. *Gilt $u \cdot v = u$, dann ist $v = \epsilon$, und falls $u \cdot v = \epsilon$, so sind $u = v = \epsilon$.*

Beweis.

(teilweise an der Tafel) □

Bemerkung (algebraische Charakterisierung)

$(\Sigma^+, \cdot)$ ist eine *Halbgruppe* und $(\Sigma^*, \cdot, \epsilon)$ ist ein *Monoid*.

Definition

Seien $A, B \subseteq \Sigma^*$ Sprachen über Σ. Die Konkatenation von A und B wird durch $A \cdot B = \{u \cdot v \mid u \in A \text{ und } v \in B\}$ beschrieben.

Für $A \subseteq \Sigma^*$ und eine nicht-negative ganze Zahl $n \in \mathbb{N}$ wird A^n induktiv definiert durch: $A^0 = \{\epsilon\}$, $A^1 = A$ sowie $A^{n+1} = A^n \cdot A$. Ferner sei $A^* = \bigcup_{n \geq 0} A^n$ und $A^+ = \bigcup_{n \geq 1} A^n$ gegeben.

Definition

Die Potenzmenge 2^A ist die Menge aller Teilmengen von A.

Beispiel ($A = \{1, 2, 3\}$)

$2^A = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$, $|2^A| = 2^{|A|} = 8$

Definition

Seien $u, v \in \Sigma^*$ Wörter über Σ. Das Wort u ist ein Präfix von $u \cdot v$ und ein Suffix von $v \cdot u$. Sind $u \neq \epsilon$ und $v \neq \epsilon$, dann ist u ein echter Präfix beziehungsweise ein echter Suffix.

Schreibweise

- u ist Präfix von v: $u \sqsubseteq v$
- u ist Suffix von v: $u \sqsupseteq v$
- u ist echter Präfix von v: $u \sqsubset v$
- u ist echter Suffix von v: $u \sqsupset v$

Lemma

Seien $u, v, w, x \in \Sigma^*$ Wörter über Σ mit $u \cdot v = w \cdot x$.

1. *Ist $|u| = |w|$, so sind $u = w$ und $v = x$.*
2. *Ist $|u| > |w|$, dann gibt es ein eindeutig bestimmtes $y \in \Sigma^+$, so daß $u = w \cdot y$ und $x = y \cdot v$ gelten. ($y \sqsubseteq x$ und $y \sqsupseteq u$)*

Definition

Eine Substitution ist eine Abbildung $\sigma : \Sigma \rightarrow 2^{\Delta^*}$, die jedem Symbol aus Σ eine Menge von Worten aus Δ^* zuordnet.

Solche Substitutionen heißen

- **endlich,** wenn für alle Symbole $x \in \Sigma$ die entsprechenden Bildmengen $\sigma(x)$ endlich sind.
- **eindeutig,** falls für alle Symbole $x \in \Sigma$ die entsprechenden Bildmengen $\sigma(x)$ nur genau ein Wort enthalten.
- **löschend,** wenn das leere Wort ϵ in mindestens einer Bildmenge vorhanden ist.

Definition

Das Bild $\sigma(w)$ eines Wortes $w = (w_1, w_2, \ldots, w_n)$ aus Σ^* unter der Substitution σ ist die Sprache $\sigma(w_1) \cdot \sigma(w_2) \cdot \cdots \cdot \sigma(w_n) \subseteq \Delta^*$.

Definition

Die Abbildung $\varphi : \Sigma^* \to \Delta^*$ ist ein (Monoid-)Homomorphismus, falls

1. für alle $u, v \in \Sigma^*$ die Eigenschaft $\varphi(u \cdot v) = \varphi(u) \cdot \varphi(v)$ gilt, und
2. das leere Wort bzgl. φ unverändert bleibt, d. h. $\varphi(\epsilon) = \epsilon$.

Bemerkung

Wir können jede *eindeutige Substitution* auch als einen solchen (Monoid-)Homomorphismus auffassen, weil die Substitution nur zeichenweise definiert wurde und die Konkatenation assoziativ ist.

Beispiel ($\Sigma = \Delta = \{\mathrm{a}, \mathrm{b}\}$)

φ mit a $\mapsto \mathrm{abb}$, b $\mapsto \mathrm{ba}$ ist ein (Monoid-)Homomorphismus.
σ mit a $\mapsto \{\mathrm{aa}, \mathrm{b}\}$, b $\mapsto \{\mathrm{b}\}$ ist *kein* Homomorphismus. (Warum?)

Schreibweise (Komposition/Iteration von Abbildungen)

Sei $\varphi : \Sigma^* \to \Sigma^*$ ein Endomorphismus, $n \in \mathbb{N}$ und $w \in \Sigma^*$ ein Wort.

$$\varphi^n(w) = (\underbrace{\varphi \circ \varphi \circ \cdots \circ \varphi}_{n\text{-mal}})(w) = \underbrace{\varphi(\varphi(\cdots \varphi}_{n\text{-mal}}(w) \cdots))$$

Beispiele (φ mit $\mathrm{a} \mapsto \mathrm{abb}, \mathrm{b} \mapsto \mathrm{ba}$)

Buchstabenweise Anwendung eines Homomorphismus:

- $\varphi(\mathrm{aba}) = \varphi(\mathrm{a}) \cdot \varphi(\mathrm{b}) \cdot \varphi(\mathrm{a}) = \mathrm{abb} \cdot \mathrm{ba} \cdot \mathrm{abb} = \mathrm{abbbaabb}$

Komposition/Iteration eines Endomorphismus:

- $\varphi^3(\mathrm{b}) = \varphi(\varphi(\varphi(\mathrm{b}))) = \varphi(\varphi(\mathrm{ba})) = \varphi(\varphi(\mathrm{b}) \cdot \varphi(\mathrm{a})) = \varphi(\mathrm{ba} \cdot \mathrm{abb}) = \mathrm{baabbabbbaba}$

Definition

Ein unendliches Wort über Σ ist eine Abbildung $w : \mathbb{N} \to \Sigma$.

Schreibweise

$w = w_0 w_1 w_2 \ldots w_i \ldots$ mit $w_i = w(i) \in \Sigma$ und Position $i \in \mathbb{N}$.

Definition

Der Präfix der Länge $k \geq 0$ des Wortes w ist $w^{[k]} = w_0 w_1 \ldots w_{k-1}$.

Ein Faktor von w ist ein Wort aus Σ^*, das in w als Teilstück enthalten ist, d. h. $u \in \Sigma^n$ ist ein Faktor (der Länge n) von w genau dann, wenn es ein $k \geq 0$ gibt mit $w^{[k+n]} = w^{[k]} \cdot u$.

Definition (Grenzwert eines unendlichen Wortes)

Sei $w_0, w_1, \ldots, w_i, \ldots$ eine Wortfolge über Σ mit der Eigenschaft, daß für alle $i \geq 1$ das Wort w_{i-1} ein echter Präfix von w_i ist.
Der Limes der Folge $(w_i)_{i \in \mathbb{N}}$ ist das unendliche Wort $u : \mathbb{N} \to \Sigma$, das für alle $i \geq 0$ die Bedingung $u^{[k]} = w_i$ erfüllt, wobei $k = |w_i|$ ist.

Schreibweise

$u = \lim_{i \to \infty} w_i$

Bemerkung (Eindeutigkeit)

Da für alle $i \geq 1$ das Wort w_{i-1} ein echter Präfix von w_i ist, sind auch für alle $0 \leq m \leq n$ die Worte w_m echte Präfixe von w_n.
Folglich ist das unendliche Wort $u = \lim_{i \to \infty} w_i$ eindeutig definiert.

Bemerkung (Existenz)

Sei $\varphi : \Sigma^* \to \Sigma^*$ ein Endomorphismus, der folgendes erfüllt:

1. $\varphi(z) \neq \epsilon$ für alle $z \in \Sigma$ (nicht-löschend), und
2. es gibt ein $u_0 \in \Sigma$ (Startsymbol) mit $\varphi(u_0) = u_0 \cdot v$ für ein $v \in \Sigma^+$.

Dann gilt $\forall n \geq 0 : \varphi^{n+1}(u_0) = \varphi^n(\varphi(u_0)) = \varphi^n(u_0 \cdot v) = \varphi^n(u_0) \cdot \varphi^n(v)$. Also ist für alle $n \geq 0$ das Wort $\varphi^n(u_0)$ ein echter Präfix des Wortes $\varphi^{n+1}(u_0)$. Folglich existiert der Limes der Folge $(\varphi^n(u_0))_{n \in \mathbb{N}}$.

Beispiel (φ mit $\mathrm{a} \mapsto \mathrm{ba}, \mathrm{b} \mapsto \mathrm{bab}$)

Der Limes existiert zwar für $(\varphi^n(\mathrm{b}))_{n \in \mathbb{N}}$ aber nicht für $(\varphi^n(\mathrm{a}))_{n \in \mathbb{N}}$.

Definition

Erfüllen φ und $u_0 \in \Sigma$ obige Existenzbedingungen, dann bezeichnen wir den Limes der Folge $(\varphi^n(u_0))_{n \in \mathbb{N}}$ als das unendliche Wort $\varphi^\omega(u_0)$, welches durch Iteration von φ aus u_0 entsteht.

Definition

Sei P eine Eigenschaft von Worten, und sei L_P die Menge der Worte, die diese Eigenschaft besitzen. P heißt faktorvererblich, wenn für jedes Wort $u \in L_P$ auch alle Faktoren von u in L_P sind.

Definition

Sei Σ ein Alphabet und $w, x \in \Sigma^*$ beliebige Worte:

1. Das Wort w heißt kubik-frei, wenn es keinen Faktor der Form xxx mit $x \neq \epsilon$ enthält.
2. Das Wort w heißt stark kubik-frei, wenn es keinen Faktor der Form $axaxa$ enthält, wobei $a \in \Sigma$ ein beliebiges Symbol ist.
3. Das Wort w heißt quadrat-frei, wenn es keinen Faktor der Form xx mit $x \neq \epsilon$ enthält.

Bemerkung: Alle drei Eigenschaften sind *faktorvererblich*.

Satz

Sei Σ ein binäres Alphabet. Jedes Wort $w \in \Sigma^*$ mit $|w| \geq 4$ enthält ein Quadrat, d. h. einen Faktor der Form xx mit $x \in \Sigma^+$.

Beweis.

Wegen der Faktorvererblichkeit der Quadratfreiheit reicht es zu prüfen, daß alle Worte der Länge 4 über Σ ein Quadrat enthalten.

$$0 \rightsquigarrow \underline{00} \qquad 0 \rightsquigarrow 01$$
$$01 \rightsquigarrow 0\underline{11} \qquad 01 \rightsquigarrow 010$$
$$010 \rightsquigarrow 01\underline{00} \qquad 010 \rightsquigarrow \underline{0101}$$

$$1 \rightsquigarrow \underline{11} \qquad 1 \rightsquigarrow 10$$
$$10 \rightsquigarrow 1\underline{00} \qquad 10 \rightsquigarrow 101$$
$$101 \rightsquigarrow 10\underline{11} \qquad 101 \rightsquigarrow \underline{1010}$$

$\square$

Die Eigenschaft, Quadrate zu enthalten, ist also für unendliche Worte über einem binären Alphabet *unvermeidbar*.

Zu Beginn des letzten Jahrhunderts (1906, 1921) beschäftigte sich der norwegische Mathematiker AXEL THUE u. a. mit folgendem unendlichen Wort. Unabhängig hat später auch MARSTON MORSE dieses Wort gefunden und untersucht (1921, 1938).

Definition

Sei $\Sigma = \{\mathrm{a}, \mathrm{b}\}$ und $\tau : \Sigma^* \to \Sigma^*$ der durch $\mathrm{a} \mapsto \mathrm{ab}, \mathrm{b} \mapsto \mathrm{ba}$ gegebene Homomorphismus. Dann ist $\tau^\omega(\mathrm{a}) = \mathrm{abbabaabbaaba}\ldots$ das unendliche Wort von Thue-Morse.

Offensichtlich enthält das Wort von Thue-Morse Quadrate. Wir werden im folgenden jedoch zeigen, daß es *stark kubik-frei* ist, d. h. es enthält *keinen* Faktor der Form $axaxa$ mit $x \in \Sigma^*$ und $a \in \Sigma$.

Bemerkung (andere Definitionen)

1. Seien $(u_n)_{n\in\mathbb{N}}$ und $(v_n)_{n\in\mathbb{N}}$ über Σ induktiv wie folgt definiert:

$$\begin{aligned} u_0 &= \mathrm{a}, & v_0 &= \mathrm{b}, & \\ u_{n+1} &= u_n \cdot v_n, & v_{n+1} &= v_n \cdot u_n & \text{für alle } n \geq 0. \end{aligned}$$

   Dann gilt $u_n = \tau^n(\mathrm{a})$, $v_n = \tau^n(\mathrm{b})$ und folglich $\lim_{i\to\infty} u_i = \tau^\omega(\mathrm{a})$.

2. Sei $\tau^\omega(\mathrm{a}) = t_0 t_1 \ldots t_n \ldots$ mit $t_i \in \Sigma$ die Darstellung des Wortes von Thue-Morse in einzelnen Buchstaben und sei $d_2(n)$ die Anzahl der Einsen in der Binärdarstellung von $n \in \mathbb{N}$. Dann haben wir folgende Aussage:

$$t_n = \begin{cases} \mathrm{a}, & \text{falls } d_2(n) \text{ gerade ist,} \\ \mathrm{b}, & \text{falls } d_2(n) \text{ ungerade ist.} \end{cases}$$

Lemma

Sei $X = \{ab, ba\}$. Für alle $w \in X^*$ gilt $awa \notin X^*$ und $bwb \notin X^*$.

Beweis.

Offensichtlich gilt für alle $x \in X^*$ die Gleichung $|x|_{\mathrm{a}} = |x|_{\mathrm{b}}$, also folgt $|w|_{\mathrm{a}} = |w|_{\mathrm{b}}$ sofort. Damit haben wir aber $|\mathrm{a}w\mathrm{a}|_{\mathrm{a}} > |\mathrm{a}w\mathrm{a}|_{\mathrm{b}}$ und $|\mathrm{b}w\mathrm{b}|_{\mathrm{b}} > |\mathrm{b}w\mathrm{b}|_{\mathrm{a}}$, was $\mathrm{a}w\mathrm{a} \notin X^*$ bzw. $\mathrm{b}w\mathrm{b} \notin X^*$ impliziert. □

Lemma

Sei $w \in \Sigma^+$ stark kubik-frei. Dann ist auch $\tau(w)$ stark kubik-frei.

Beweis.

(an der Tafel) □

Satz

Das unendliche Wort von Thue-Morse ist stark kubik-frei.

Beweis.

Der Homomorphismus τ vererbt die Eigenschaft, stark kubik-frei zu sein (siehe letztes Lemma). Da das Wort a stark kubik-frei ist, sind auch die Worte $\tau^n(\mathrm{a})$ für alle $n \geq 0$ stark kubik-frei. Damit enthält das unendliche Wort $\tau^\omega(\mathrm{a})$ ebenfalls keinen Faktor der Form *cxcxc* mit $c \in \Sigma$ und $x \in \Sigma^*$. $\square$

Folglich ist auch die Eigenschaft, Muster der Form *xxx* (also Kuben) zu enthalten, über einem binären Alphabet *vermeidbar*.

Fragestellung: Gibt es unendliche quadrat-freie Worte?

Wir haben bereits gezeigt, daß diese Frage für binäre Alphabete negativ zu beantworten ist. Aber existieren vielleicht unendliche quadrat-freie Worte für Alphabete mit mehr als zwei Symbolen?

Seien $\Sigma = \{\mathsf{a}, \mathsf{b}\}$ und $\Gamma = \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$ zwei endliche Alphabete.

Definition

Sei $\mu : \Gamma^* \to \Gamma^*$ der durch $\mathtt{0} \mapsto \mathtt{012}$, $\mathtt{1} \mapsto \mathtt{02}$, $\mathtt{2} \mapsto \mathtt{1}$ beschriebene Homomorphismus. Dann ist $\mu^\omega(\mathtt{0}) = \mathtt{012021012102012021 02}\ldots$ das unendliche Wort von Morse.

Wir werden nun zeigen, daß das unendliche Wort von Morse $\mu^\omega(\mathtt{0})$ quadrat-frei ist. Dafür benutzen wir einen weiteren Homomorphismus, der eine Beziehung zum unendlichen Wort von Thue-Morse herstellt.

Definition

Der Homomorphismus $\delta : \Gamma^* \to \Sigma^*$ sei durch die Abbildungsregeln $\mathtt{0} \mapsto \mathtt{abb}$, $\mathtt{1} \mapsto \mathtt{ab}$, $\mathtt{2} \mapsto \mathtt{a}$ definiert.

Lemma

Für alle $n \geq 0$ gilt $\tau^n(\mathtt{a}) = \delta(\mu^n(\mathtt{0}))$ und folglich $\tau^\omega(\mathtt{a}) = \delta(\mu^\omega(\mathtt{0}))$, d. h. der Homomorphismus δ bildet das unendliche Wort von Morse auf das unendliche Wort von Thue-Morse ab.

Beweisskizze

$$\begin{array}{ccccccccccc}
\mathtt{0} & \xrightarrow{\mu} & \mathtt{012} & \qquad & \mathtt{1} & \xrightarrow{\mu} & \mathtt{02} & \qquad & \mathtt{2} & \xrightarrow{\mu} & \mathtt{1} \\
\delta\downarrow & & \delta\downarrow & & \delta\downarrow & & \delta\downarrow & & \delta\downarrow & & \delta\downarrow \\
\mathtt{abb} & \xrightarrow{\tau} & \mathtt{abbaba} & & \mathtt{ab} & \xrightarrow{\tau} & \mathtt{abba} & & \mathtt{a} & \xrightarrow{\tau} & \mathtt{ab}
\end{array}$$

Satz

Das unendliche Wort von Morse ist quadrat-frei.

Beweis.

Angenommen, $\mu^\omega(\mathtt{0})$ enthält ein Quadrat. Folglich hat es einen Präfix der Form $uxxv$ mit $u \in \Gamma^*$, $x \in \Gamma^+$ und $v \in \Gamma$. Nach dem vorherigen Lemma ist $\delta(uxxv) = \delta(u)\delta(x)\delta(x)\delta(v)$ auch ein Präfix von $\tau^\omega(\mathrm{a})$. Nun gilt aber $\delta(x) = \mathrm{a}y$ und $\delta(v) = \mathrm{a}w$ für gewisse $y, w \in \Sigma^*$, weil die Abbildung δ immer ein a am Anfang erzeugt.

Damit hätte das unendliche Wort von Thue-Morse einen Präfix der Form $\delta(u)\mathrm{a}y\mathrm{a}y\mathrm{a}w$, was jedoch ein Widerspruch zu seiner starken Kubikfreiheit wäre. Also ist $\mu^\omega(\mathtt{0})$ quadrat-frei. □

Die Eigenschaft, Quadrate zu enthalten, ist für unendliche Worte über einem Alphabet mit mehr als zwei Buchstaben *vermeidbar*.

Definition

Die Fibonacci-Folge wird induktiv durch folgende Regeln definiert:

$$F_0 = 0, \qquad F_1 = 1, \qquad F_n = F_{n-1} + F_{n-2}$$

Definition

Das Fibonacci-Wort $\phi^\omega(\mathrm{a}) = \mathrm{abaababaabaababaababaabaab}\ldots$ wird durch den Homomorphismus ϕ mit $\mathrm{a} \mapsto \mathrm{ab}$, $\mathrm{b} \mapsto \mathrm{a}$ beschrieben.

Fakt

Für $k \in \mathbb{N}$ gilt $|\phi^k(a)| = F_{k+2}$, $|\phi^k(a)|_a = F_{k+1}$ und $|\phi^k(a)|_b = F_k$.

Fakt

Das Fibonacci-Wort ist zwar biquadrat-frei jedoch nicht kubik-frei.

Definition

Die Zimin-Muster werden induktiv beschrieben durch

$$Z_0 = \epsilon, \qquad Z_{n+1} = Z_n x_{n+1} Z_n \,.$$

Beispiele

- $Z_1 = x_1$
- $Z_2 = x_1 x_2 x_1$
- $Z_3 = x_1 x_2 x_1 x_3 x_1 x_2 x_1$

Fakt

Alle Zimin-Muster sind für beliebige Alphabete unvermeidbar.

- Einleitung
  - Kombinatorik auf Wörtern, Querverbindungen
  - Operationen (Konkatenation), Homomorphismen
  - Unendliche Wörter als Iteration von Endomorphismen
- Vermeidbare und unvermeidbare Muster
  - Wort von Thue-Morse
  - Wort von Morse (unendliches quadrat-freies Wort)
  - Fibonacci-Wort und Zimin-Muster
- Ausblick
  - Beispiel: (Satz von VAN DER WAERDEN)
    Für „hinreichend" lange Worte tritt ein Symbol/Buchstabe „beliebig" oft in „festen Abständen" auf.
  - Anwendung: Ausnutzung solcher unvermeidbarer Regelmäßigkeiten beim Algorithmenentwurf.
  - Weitere Beispiele: (mit Zusammenhang zur „Kombinatorik auf Wörtern")
    Hilberts flächenfüllende Kurve, Cantor-Staub (Fraktal), . . .

M. Lothaire (Autorenkollektiv)
Combinatorics on Words
Encyclopedia of Mathematics 17, Addison-Wesley, 1983.

M. Lothaire (Autorenkollektiv)
Algebraic Combinatorics on Words
Encyclopedia of Mathematics 90, Cambridge University Press, 2002.
`http://www-igm.univ-mlv.fr/~berstel/Lothaire/`

Friedrich Otto
Worte und ihre kombinatorischen Eigenschaften
Vorlesung WS 1995/96, Universität Kassel.

Johannes Waldmann
L-Systeme
Vorlesung SS 2001, Universität Leipzig.

Juhani Karhumäki
Combinatorics on Words: A New Challenging Topic
TUCS Technical Report, No. 645, 2004.