Avoidable Words

George F. McNulty

University of South Carolina

Columbia, SC 29212

USA

mcnulty@math.sc.edu

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.
- Computer Science.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.
- Computer Science.
- Automata and Formal Languages.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.
- Computer Science.
- Automata and Formal Languages.
- Algebra.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.
- Computer Science.
- Automata and Formal Languages.
- Algebra.
- Topology.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.
- Computer Science.
- Automata and Formal Languages.
- Algebra.
- Topology.
- Number Theory.

Strings of Symbols and Where They Might Come Up

Mathematicians are in the business of producing strings of symbols. In some parts of mathematics, strings of symbols themselves become the objects of mathematical attention.

- Mathematical Logic.
- Computer Science.
- Automata and Formal Languages.
- Algebra.
- Topology.
- Number Theory.
- Dynamical Systems.

Patterns in Strings of Symbols

Most of the strings of symbols that arise in these displines display within themselves instances of patterns and other regularities.

It is the investigation of the occurrences of patterns in strings of symbols that is the topic of these lectures.

An **alphabet** is just a set of letters

An **alphabet** is just a set of letters and a **letter** is just a member of some alphabet.

An **alphabet** is just a set of letters and a **letter** is just a member of some alphabet.

Today, all alphabets are finite.

An **alphabet** is just a set of letters and a **letter** is just a member of some alphabet.

Today, all alphabets are finite.

A **word** is just a finite string of letters. For an alphabet Σ we use Σ^* to denote the set of all words over the alphabet Σ.

An **alphabet** is just a set of letters and a **letter** is just a member of some alphabet.

Today, all alphabets are finite.

A **word** is just a finite string of letters. For an alphabet Σ we use Σ^* to denote the set of all words over the alphabet Σ. One of these words is the **empty word**, which has length 0.

An **alphabet** is just a set of letters and a **letter** is just a member of some alphabet.

Today, all alphabets are finite.

A **word** is just a finite string of letters. For an alphabet Σ we use Σ^* to denote the set of all words over the alphabet Σ. One of these words is the **empty word**, which has length 0.

Σ^+ denotes the set of all nonempty words over Σ.

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

⋮

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

⋮

The Jacobson-Schutzenberger-Zimin Sequence

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

⋮

The Jacobson-Schutzenberger-Zimin Sequence

a

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

⋮

The Jacobson-Schutzenberger-Zimin Sequence

a

aba

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

$\vdots$

The Jacobson-Schutzenberger-Zimin Sequence

a

aba

abacaba

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

⋮

The Jacobson-Schutzenberger-Zimin Sequence

a

aba

abacaba

abacabadabacaba

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a
ab
abba
abbabaab
abbabaabbaababba
⋮

The Jacobson-Schutzenberger-Zimin Sequence

a
aba
abacaba
abacabadabacaba
abacabadabacabaeabacabadabacaba

Two Interesting Sequences of Words

The Prouhet-Thue-Morse Sequence

a

ab

abba

abbabaab

abbabaabbaababba

⋮

The Jacobson-Schutzenberger-Zimin Sequence

a

aba

abacaba

abacabadabacaba

abacabadabacabaeabacabadabacaba

⋮

What is a Pattern?

A pattern is, in essence, a system of specifications that can be embodied and scaled in an assortment of ways.

For example, in

abbabaabbaababba

We can see the following:

What is a Pattern?

A pattern is, in essence, a system of specifications that can be embodied and scaled in an assortment of ways.

For example, in

abbabaabbaababba

We can see the following:

abbabaabbaababba

What is a Pattern?

A pattern is, in essence, a system of specifications that can be embodied and scaled in an assortment of ways.

For example, in

abbabaabbaababba

We can see the following:

abbabaabbaababba

abbabaabbaababba

What is a Pattern?

A pattern is, in essence, a system of specifications that can be embodied and scaled in an assortment of ways.

For example, in

abbabaabbaababba

We can see the following:

abbabaabbaababba

abbabaabbaababba

abbabaabbaababba

What is a Pattern?

A pattern is, in essence, a system of specifications that can be embodied and scaled in an assortment of ways.

For example, in

abbabaabbaababba

We can see the following:

abbabaabbaababba

abbabaabbaababba

abbabaabbaababba

All the red words are instances of a single pattern: the pattern

xx

So patterns like xx are just words.

What is a Pattern?

A pattern is, in essence, a system of specifications that can be embodied and scaled in an assortment of ways.

For example, in

abbabaabbaababba

We can see the following:

abbabaabbaababba

abbabaabbaababba

abbabaabbaababba

All the red words are instances of a single pattern: the pattern

xx

So patterns like xx are just words.

A word W is an **instance** or an **image** of a word U, provided W can be obtained from U by systematically substituting nonempty words for the letters of U.

Thus baabbaab is the instance of xx obtained by substituting the word baab for the letter x.

As another example

$$\mathrm{xyyxyxxy}$$

is the image of

$$\mathrm{abba}$$

via

$$a \mapsto xy \quad \text{and} \quad b \mapsto yx$$

Avoiding Patterns

We say a word V is a **subword** of a word W when W=XVY for some (possibly empty) words X and Y.

Avoiding Patterns

We say a word V is a **subword** of a word W when W=XVY for some (possibly empty) words X and Y.

The word W **encounters** the word U if and only if some instance of U is a subword of W. We say W **avoids** U otherwise

Avoiding Patterns

We say a word V is a **subword** of a word W when W=XVY for some (possibly empty) words X and Y.

The word W **encounters** the word U if and only if some instance of U is a subword of W. We say W **avoids** U otherwise—that is when no instance of U is a subword of W.

Avoidable and Unavoidable Wrods

The word U is **avoidable on the alphabet** Σ provided infinitely many words in Σ^+ avoid U.

Avoidable and Unavoidable Wrods

The word U is **avoidable on the alphabet** Σ provided infinitely many words in Σ^+ avoid U. Of course, the only thing that matters is the size of Σ. So if Σ has k elements we also say that U is k**-avoidable**.

Avoidable and Unavoidable Wrods

The word U is **avoidable on the alphabet** Σ provided infinitely many words in Σ^+ avoid U. Of course, the only thing that matters is the size of Σ. So if Σ has k elements we also say that U is k**-avoidable**.

The word U is **avoidable** if and only if U is k-avoidable for some natural number k.

Avoidable and Unavoidable Wrods

The word U is **avoidable on the alphabet** Σ provided infinitely many words in Σ^+ avoid U. Of course, the only thing that matters is the size of Σ. So if Σ has k elements we also say that U is k**-avoidable**.

The word U is **avoidable** if and only if U is k-avoidable for some natural number k.Words that fail to be avoidable are said to be **unavoidable**.

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
	a	
	ab	
	aa	
	aba	
	aaa	
	abacaba	

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
	ab	
	aa	
	aba	
	aaa	
	abacaba	

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
ab		
	aa	
	aba	
	aaa	
	abacaba	

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
ab		
		aa Axel Thue 1906
	aba	
	aaa	
	abacaba	

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
ab		
		aa Axel Thue 1906
aba		
	aaa	
	abacaba	

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
ab		
		aa Axel Thue 1906
aba		
		aaa
	abacaba	

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
ab		
		aa Axel Thue 1906
aba		
		aaa
abacaba		

Are Any Words Avoidable? Are Any Words Unavoidable?

There better be some words of each kind!

Unavoidable	Word	Avoidable
a		
ab		
		aa Axel Thue 1906
aba		
		aaa
abacaba		

Indeed all the Zimin words, as well as any word encountered by a Zimin word, can be easily seen to be unavoidable.

The Adjacency Graph of a Word

The **adjacency graph** of a word W is a bipartite graph whose vertices are divided into two parts: a left copy of the alphabet and a right copy of the alphabet. The edges are determined by scanning the word. There is an edge joining the left copy of the letter x to the right copy of the letter y if and only if the word xy is a subword of the word W.

An example will serve to show how this works.

The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

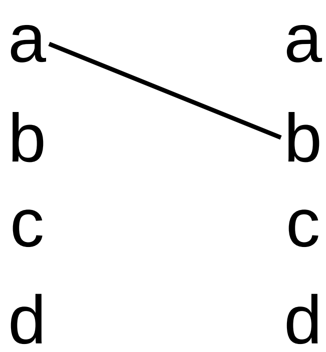
abacabadabacaba

a	a
b	b
c	c
d	d

The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

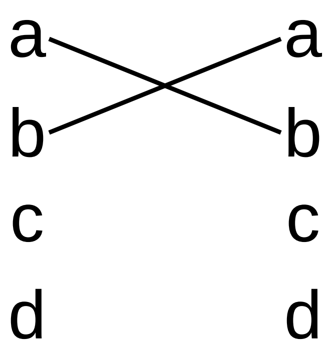
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

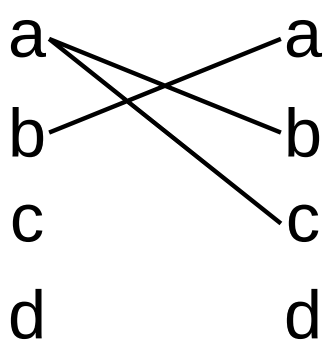
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

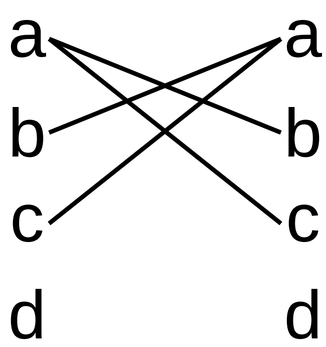
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

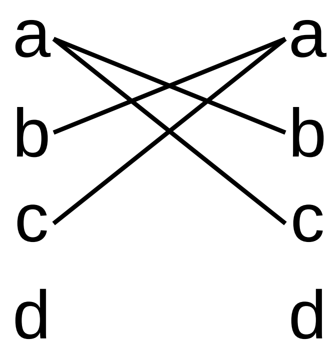
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

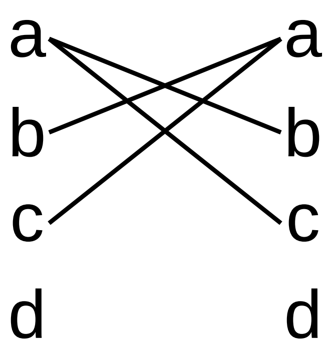
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

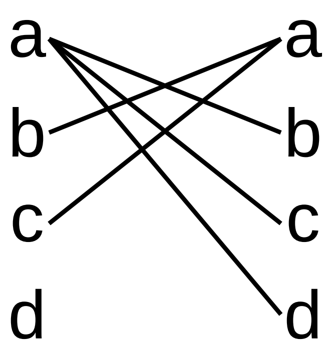
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

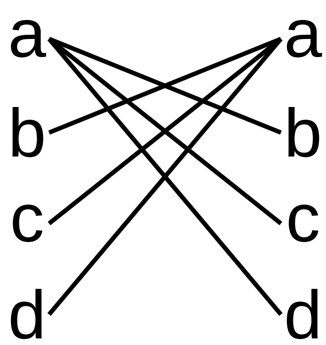
abacabadabacaba



The Adjacency Graph of abacaba

As we slide a window of length 2 along abacabadabacaba we insert edges as shown.

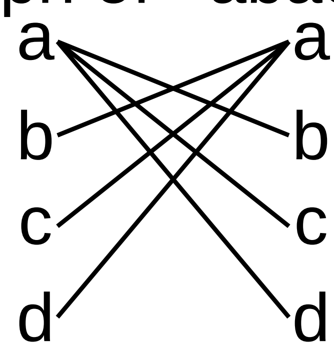
abacabadabacaba



Free Sets

For a word W on an alphabet Σ we say that $F \subseteq \Sigma$ is **free** for W when no connected component of the adjacency graph of W contains both an element of the left copy of F and the right copy of F.

Since the adjacency graph of abacabadabacaba is



we see that the free sets are just all the subsets of these sets: $\{$a$\}$ and $\{$b,c,d$\}$.

Deletion of Letters from Words

Suppose W is a word and F is a set of letters. We can delete the letters in F from W by just erasing them. This leaves gaps, but by pushing the result together we get another word.

Let's delete $\{$a$\}$ from

abacabadabacaba

Deletion of Letters from Words

Suppose W is a word and F is a set of letters. We can delete the letters in F from W by just erasing them. This leaves gaps, but by pushing the result together we get another word.

Let's delete $\{$a$\}$ from

abacabadabacaba
b c b d b c b

Deletion of Letters from Words

Suppose W is a word and F is a set of letters. We can delete the letters in F from W by just erasing them. This leaves gaps, but by pushing the result together we get another word.

Let's delete $\{$a$\}$ from

abacabadabacaba
bcbdbcb

Reducing Words by Deleting Free Sets

We say a word W **reduces in one step** to a word U provided U can be obtained from W by deleting all the letters in some set free for W.

Reducing Words by Deleting Free Sets

We say a word W **reduces in one step** to a word U provided U can be obtained from W by deleting all the letters in some set free for W. So abacabadabacaba reduces in one step to bcbdbcb.

Reducing Words by Deleting Free Sets

We say a word W **reduces in one step** to a word U provided U can be obtained from W by deleting all the letters in some set free for W. So $abacabadabacaba$ reduces in one step to $bcbdbcb$.

We say a word W **reduces** to a word U if and only if U can be obtained from W by a sequence of one-step reductions.

A Sample Reduction

Now

- a is free for abacabadabacaba
- b is free for bcbdbcb
- c is free for cdc and finally
- d is free for d.

So the following reduction is available.

abacabadabacaba

A Sample Reduction

Now

- a is free for abacabadabacaba
- b is free for bcbdbcb
- c is free for cdc and finally
- d is free for d.

So the following reduction is available.

abacabadabacaba
b c b d b c b

A Sample Reduction

Now

- a is free for abacabadabacaba
- b is free for bcbdbcb
- c is free for cdc and finally
- d is free for d.

So the following reduction is available.

abacabadabacaba

bcbdbcb

A Sample Reduction

Now

- a is free for abacabadabacaba
- b is free for bcbdbcb
- c is free for cdc and finally
- d is free for d.

So the following reduction is available.

abacabadabacaba
bcbdbcb
c d c

A Sample Reduction

Now

- a is free for abacabadabacaba
- b is free for bcbdbcb
- c is free for cdc and finally
- d is free for d.

So the following reduction is available.

abacabadabacaba
bcbdbcb
cdc

A Sample Reduction

Now

- a is free for abacabadabacaba
- b is free for bcbdbcb
- c is free for cdc and finally
- d is free for d.

So the following reduction is available.

abacabadabacaba
bcbdbcb
cdc
d

A Characterization of Unavoidability

Theorem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982). *Let W be a word. The following are equivalent:*

- *W is unavoidable.*

A Characterization of Unavoidability

Theorem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982). *Let W be a word. The following are equivalent:*

- *W is unavoidable.*
- *W can be reduced to the empty word.*

A Characterization of Unavoidability

Theorem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982). *Let W be a word. The following are equivalent:*

- *W is unavoidable.*
- *W can be reduced to the empty word.*
- *The Zimin word on α letters encounters W, where α is the number of distinct letters in W.*

A Characterization of Unavoidability

Theorem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982). *Let W be a word. The following are equivalent:*

- *W is unavoidable.*
- *W can be reduced to the empty word.*
- *The Zimin word on α letters encounters W, where α is the number of distinct letters in W.*

This theorem arose in the late 1970's in the work of Dwight Bean, Andrzej Ehrenfeucht, and George McNulty and independently in the work of I. A. Zimin. The last condition is due solely to Zimin.

Consequences of Unavoidability

If W is unavoidable, then some letter occurs exactly once in W.

Consequences of Unavoidability

If W is unavoidable, then some letter occurs exactly once in W.

FACT: Every word on an n letter alphabet which has length at least 2^n has a subword in which every letter that occurs occurs at least twice.

Consequences of Unavoidability

If W is unavoidable, then some letter occurs exactly once in W.

FACT: Every word on an n letter alphabet which has length at least 2^n has a subword in which every letter that occurs occurs at least twice.

Call a word **long** if it has length at least 2^n where n is the number of distinct letters occurring in the word. We see that every long word is avoidable. The word xx is long even though it looks short. We have at least part of Thue's findings: xx is avoidable.

The Zimin word on n letters has length $2^n - 1$ and it is, as we noted, unavoidable. Indeed, it is as long as an unavoidable word on the n-letter alphabet can be.

The Zimin word on n letters has length $2^n - 1$ and it is, as we noted, unavoidable. Indeed, it is as long as an unavoidable word on the n-letter alphabet can be.

So for word to be unavoidable it must be of rather modest length.

The Zimin word on n letters has length $2^n - 1$ and it is, as we noted, unavoidable. Indeed, it is as long as an unavoidable word on the n-letter alphabet can be.

So for word to be unavoidable it must be of rather modest length.

On the n-letter alphabet there are only finitely many unavoidable words. An industrious mathematician might try to describe them all. Indeed, one might go on to discover how many letters it takes to avoid the avoidable ones. After some preliminary work by U. Schmidt and P. Roth, J. Cassaigne managed to classify all words on 2 letters. As far as I know, nobody knows a complete description for words on 3 letters.

Today's Open Problem

What is the computational complexity of the set of unavoidable words?

A Short Side Trip: Squarefree Words

A word W is **squarefree** provided it avoids the pattern xx. Axel Thue proved in 1906 that there are infinitely many squarefree words on the 3 letter alphabet. Over the course of the ensuing 70 years Thue's work was rediscovered again and again. Since the mid-1970's many interesting things have been found out about squarefreeness. A look at these results can give some indication of what directions the study of avoidable words more generally might take.

How to Make Infinitely Many Squarefree Words

Theorem (Axel Thue, 1912). *Let* Σ *and* Γ *be alphabets and let* h *be a homomorphism from* Σ^+ *into* Γ^+*. If*

How to Make Infinitely Many Squarefree Words

Theorem (Axel Thue, 1912). *Let Σ and Γ be alphabets and let h be a homomorphism from Σ^+ into Γ^+. If*

- $h(W)$ *is squarefree whenever* W *is a squarefee word of length at most* 3, *and*

How to Make Infinitely Many Squarefree Words

Theorem (Axel Thue, 1912). *Let Σ and Γ be alphabets and let h be a homomorphism from Σ^+ into Γ^+. If*

- $h(W)$ *is squarefree whenever W is a squarefee word of length at most* 3*, and*
- $a = b$ *whenever* $a, b \in \Sigma$ *such that* $h(a)$ *is a submword of* $h(b)$*,*

How to Make Infinitely Many Squarefree Words

Theorem (Axel Thue, 1912). *Let Σ and Γ be alphabets and let h be a homomorphism from Σ^+ into Γ^+. If*

- $h(W)$ *is squarefree whenever* W *is a squarefee word of length at most* 3*, and*
- $a = b$ *whenever* $a, b \in \Sigma$ *such that* $h(a)$ *is a submword of* $h(b)$*,*

then $h(U)$ *is squarefree whenever* $U \in \Sigma^+$ *is squarefree.*

Some Squarefree Homomorphisms

$$a \mapsto abcab$$

$$b \mapsto acabcb$$

$$c \mapsto acbcacb$$

Axel Thue, 1912

Some Squarefree Homomorphisms

$$a \mapsto abacb$$

$$b \mapsto abcbac$$

$$c \mapsto abcacbc$$

Axel Thue, 1912

Some Squarefree Homomorphisms

$$a \mapsto abcbacbcabcba$$

$$b \mapsto bcacbacabcacb$$

$$c \mapsto cabacbabcabac$$

Jonathan Leech, 1957

Some Squarefree Homomorphisms

$$
\begin{aligned}
a &\mapsto abcd \\
b &\mapsto abdc \\
c &\mapsto acbd \\
d &\mapsto acdb \\
e &\mapsto adbc
\end{aligned}
$$

Bean, Ehrenfeucht, McNulty, 1979

Some Squarefree Homomorphisms

$$
\begin{aligned}
a_0 &\mapsto dW_0eW_0 \\
a_1 &\mapsto dW_1eW_1 \\
a_2 &\mapsto dW_2eW_2 \\
&\vdots \qquad \vdots
\end{aligned}
$$

The words $W_0, W_1, W_2, \ldots$ is just the infinite list of squarefree words of $\{a, b, c\}$.

Bean, Ehrenfeucht, McNulty, 1979

Some Squarefree Homomorphisms

There is even a squarefree homomorphism from words on 4 letters to words on 3 letters. But you don't want to see it. The images of each of the three letters are words of length about 200.

Theorem (Bean, Ehrenfeucht, McNulty 1979). *Let Σ be the countably infinite alphabet and let Γ be the three-letter alphabet. There is a squarefree homomorphism from Σ^+ into Γ^+.*

Ordering Words

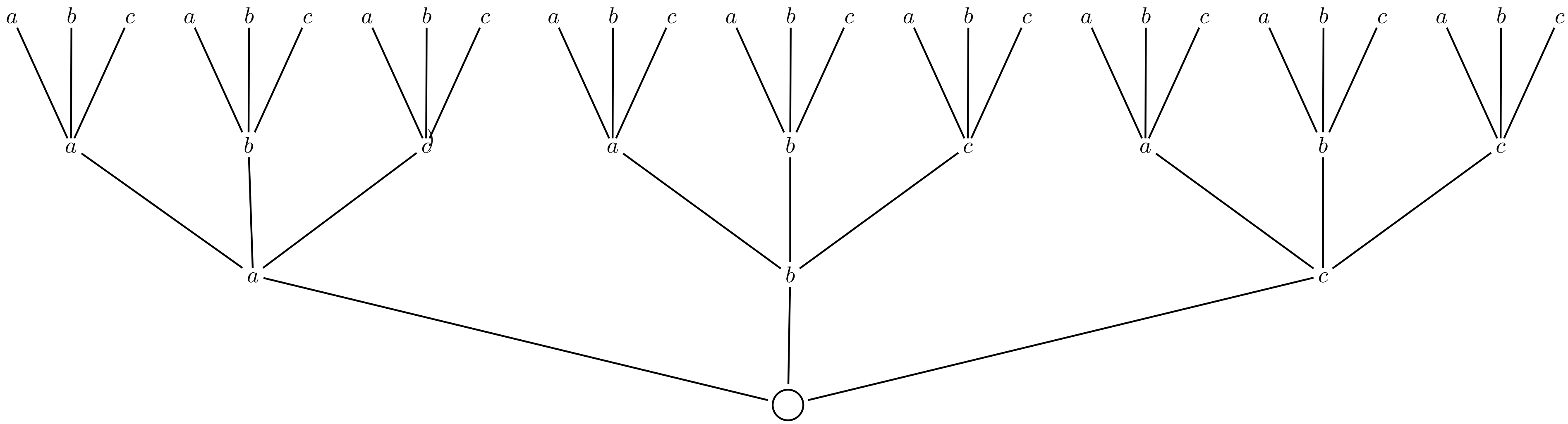
There are several interesting orders that can be imposed on sets of words. One is the ordering by initial segment. We say that the word W is an **initial segment** of the word U provided $U = WV$ for some, possibly empty, word U. Let us write $W \leq U$ when W is an initial segment of U.
It is easy to see that $\leq$ is reflexive, transitive and antisymmetric. That is $\leq$ is an order.
So for any alphabet Σ we see that $\langle \Sigma^*, \leq \rangle$ is an ordered set.

The Tree of Words

One way to display $\langle \Sigma^*, \leq \rangle$ is as a (rooted, ordered)tree. Suppose $\Sigma = \{a, b, c\}$ We impose the ordinary linear (alphabetical) ordering on Σ. Then we take the full ternary tree and label the nodes with the letters following a left to right pattern. The root we leave unlabelled.

The first few levels of this tree are displayed next.

The Bottom of the Tree



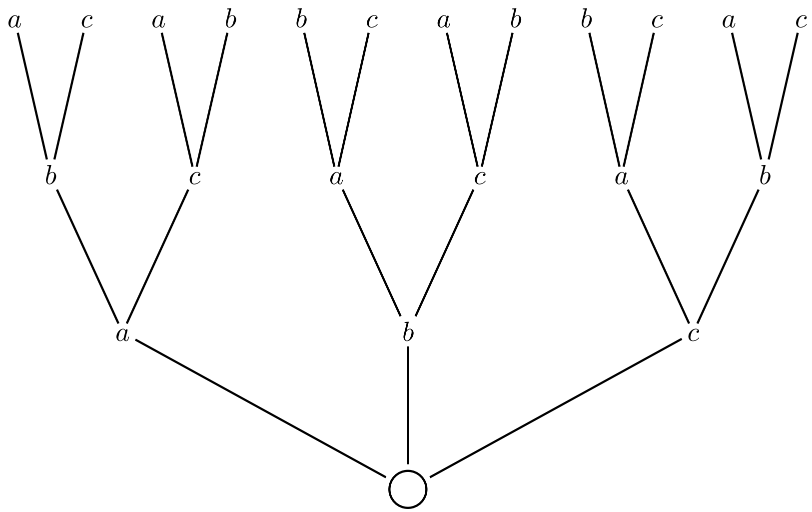
Walking Through the Tree

A walk starting at the root and passing up along the branches produces a word, letter by letter, as you pass over the branch points.

The tree has infinitely many levels. We could identify the infinite branches through the tree with the points in the unit interval (in this case expressed in ternary notation rather than decimal notation). This bring topology into the picture.

We could also agree to only include those walks along the branched which avoided some word W.

The Bottom of the Squarefree Tree



Finite Branches

Think of the last tree as a pruned version of the full ternary tree. Right now it looks like the full binary tree. But draw in a few more levels and you see that it is not.

For instance, the branch that reads $abacaba$ cannot be extended. It is a maximal squarefree word.

The number $g(n)$ of nodes at level n is the number of squarefree words of length n. The function g is the growth function for words avoiding xx.

We could do the same for every word W in place of xx.

Theorem (Bean, Ehrenfeucht, McNulty 1979). *Let W be a k^{th}-powerfree word on the finite alphabet Σ. W is a subword of some maximal k^{th}-powerfree word on Σ.*

The Growth of Squarefree Words

Theorem (Brandenburg 1983; Brinkhuis 1983). *On the three-letter alphabet, the growth rate of squarefree words has an exponential lower bound.*

In the Space of Infinite Words. . .

Theorem (Shelton and Soni 1980-83). *The space of squarefree ω-words on the three letter alphabet is perfect.*

Another Order

Let $W \triangleleft U$ mean that U ancounters W. This relation on words in reflexive and transitive. It fails however to be antisymmetric. But we can fix this.

Regard to words as **literally similar** provided there is a one-to-one map of the alphabet of one word onto the alphabet of the other that carries the first word to the second.

So cat and dog are literally similar while $read$ and $reed$ are not.

Modulo literal similarity $\triangleleft$ is an ordering on words.

The Ordered Set of Unavoidable Words

Let Σ be a finite alphabet. The set Γ of unavoidable words is Σ^* is finite. Up to literal similarity $\langle \Gamma, \lhd \rangle$ is a finite ordered set. It would be interesting to know something about its structure.

Today's Open Problem

For each natural number n find out how many unavoidable words there are on the n letter alphabet

How Many Letters Does It Take?

Let W be any word. We use $\alpha(\mathsf{W})$ to denote the number of distinct letters appearing in W.

If W is avoidable, then there is a smallest natural number $\mu(\mathsf{W})$ such that there is an infinite word on μ letters which avoids W.

We can extend this function μ to all words by putting $\mu(\mathsf{W}) = \infty$ when W is unavoidable.

How Many Letters Does It Take?

Let W be any word. We use $\alpha(\mathsf{W})$ to denote the number of distinct letters appearing in W.

If W is avoidable, then there is a smallest natural number $\mu(\mathsf{W})$ such that there is an infinite word on μ letters which avoids W.

We can extend this function μ to all words by putting $\mu(\mathsf{W}) = \infty$ when W is unavoidable.

PROBLEM: Is μ a recursive function? That is, is there are computer program which computes μ?

How Many Letters Does It Take?

Let W be any word. We use $\alpha(\mathsf{W})$ to denote the number of distinct letters appearing in W.

If W is avoidable, then there is a smallest natural number $\mu(\mathsf{W})$ such that there is an infinite word on μ letters which avoids W.

We can extend this function μ to all words by putting $\mu(\mathsf{W}) = \infty$ when W is unavoidable.

PROBLEM: Is μ a recursive function? That is, is there are computer program which computes μ?

PROBLEM: What is the assymptotic behaviour of μ with respect to α?

The Word $\Omega_{m,z}$

The words mentioned in the title of this slide extend infinitely to the right. The parameters m and z are natural numbers with $m \geq 2$ and $z \geq 1$. The number of distinct letters in $\Omega_{m,z}$ is $m\lceil m^{1/z}\rceil$. I am only going to show you $\Omega_{2,1}$. It's alphabet has size 4. It proves convenient to use $\{0, 1, 2, 3\}$ as the alphabet.

The Word $\Omega_{m,z}$

Let φ be the endomorphism given by

$$
\begin{aligned}
0 &\mapsto 01\\
1 &\mapsto 21\\
2 &\mapsto 03\\
3 &\mapsto 23
\end{aligned}
$$

The Word $\Omega_{m,z}$

Let φ be the endomorphism given by

$$0 \mapsto 01$$
$$1 \mapsto 21$$
$$2 \mapsto 03$$
$$3 \mapsto 23$$

Start from 0 and iterate φ:

$$0$$
$$01$$
$$0121$$
$$01210321$$
$$0121032101230321$$

Dean's Word: $\Omega_{2,1}$

$$0$$

$$01$$

$$0121$$

$$01210321$$

$$0121032101230321$$

$$\cdots$$

Dean's Word: $\Omega_{2,1}$

$$0$$

$$01$$

$$0121$$

$$01210321$$

$$0121032101230321$$

$$\cdots$$

Notice that at each stage the previous word occurs as an initial segment. (This is a consequence of the fact that the image of 0 begins with 0.) Continuing in this way we construct the infinite word $\Omega_{2,1}$.

Three Theorems

Theorem (Baker, McNulty, Taylor 1989). *For integers* r, m, z *with* $r, z > 0$ *and* $m > (r+1)z$*, the word* $\Omega_{m,z}$ *avoids all avoidable words on* r *or fewer letters.*

Three Theorems

Theorem (Baker, McNulty, Taylor 1989). *For integers r, m, z with $r, z > 0$ and $m > (r+1)z$, the word $\Omega_{m,z}$ avoids all avoidable words on r or fewer letters.*

Call a word W **locked** provided it cannot be reduced to any other word.

Theorem (Baker, McNulty, Taylor 1989). *Dean's word $\Omega_{2,1}$ avoids every locked word.*

Three Theorems

Theorem (Baker, McNulty, Taylor 1989). *For integers r, m, z with $r, z > 0$ and $m > (r+1)z$, the word $\Omega_{m,z}$ avoids all avoidable words on r or fewer letters.*

Call a word W **locked** provided it cannot be reduced to any other word.

Theorem (Baker, McNulty, Taylor 1989). *Dean's word $\Omega_{2,1}$ avoids every locked word.*

Theorem (I. Mel'nichuk 1988). $\mu < \alpha + 6$.

Walter's Word Δ

Let Δ be the word abwbcxcaybazac. It is a bit easier to understand Δ like this:

ab w bc x ca y ba z ac

Walter's Word Δ

Let Δ be the word abwbcxcaybazac. It is a bit easier to understand Δ like this:

ab w bc x ca y ba z ac

It is easy to see that Δ is a locked word. This means that it is avoided by $\Omega_{2,1}$ and so Δ is 4-avoidable.

Walter's Word Δ

Let Δ be the word abwbcxcaybazac. It is a bit easier to understand Δ like this:

ab w bc x ca y ba z ac

It is easy to see that Δ is a locked word. This means that it is avoided by $\Omega_{2,1}$ and so Δ is 4-avoidable.

Walter's word Δ is not 3-avoidable.

Walter's Word Δ

Let Δ be the word abwbcxcaybazac. It is a bit easier to understand Δ like this:

ab w bc x ca y ba z ac

It is easy to see that Δ is a locked word. This means that it is avoided by $\Omega_{2,1}$ and so Δ is 4-avoidable.

Walter's word Δ is not 3-avoidable.

Theorem (Baker, McNulty, Taylor 1989).

- *The growth rate of words on 4 letter which avoid Δ has a quadratic lower bound and a polynomial upper bound.*
- *On four letters the space of $\mathbb{Z}$-words avoiding Δ is perfect, and in fact a Cantor space.*

The Words of Ronald Clark, 2001

Δ is 4-avoidable but not 3-avoidable. That is $\mu(\Delta) = 4$. Let rho be the following word:

ab u ba w ac x bc y cda z dcd

Ronald Clark, in his Ph.D. dissertation written at UCLA in 2001 under the direction of Kirby Baker, proved that rho is 5-avoidable but not 4 avoidable.

Some Applications

Burnside's conjecture was that every finitely generated group of finite exponent is finite. This conjecture originated in 1905 and was not resolved (in the negative) until the work of Adjan and Novikov in 1968. The Adjan Novikov construction is genuinely elaborate, but it relies on the avoidability of xx and xxx.

Some Applications

Burris and Nelson (1971), and Jarda Jezek (1976) used the avoidability of xx to show that lattice of subvarieties of the variety of semigroups must have intervals of very high complexity. In fact Jezek produces an infinite set $\mathcal{F}$ of squarefree words on the three letter alphabet so that $\mathcal{F} - \{W\}$ avoids W for every $W \in \mathcal{F}$.

Some Applications

Mark Sapir (1987) used the full weight of avoidable words to give an alogithm for determining which finite semigroups have inherently nonfinitely based equational theories.

Some Applications

With the help of the theory of avoidability, Margolis and Sapir (1995) proved that no finite semigroup can be inherently nonfinitely Q-based.

Some Applications

A recent unpublished result of Ralph McKenzie is that given any function f on the natural numbers there is a finite semigroup $\mathbf{S}$ So that the free spectrum function of $\mathbf{S}$ dominates f.

Open Problems

PROBLEM 0: What is the computational complexity of the set of avoidable words?

Open Problems

PROBLEM 0: What is the computational complexity of the set of avoidable words?

PROBLEM 1: How many unavoidable words are there on the alphabet with n letters? Describe these finite ordered sets.

Open Problems

PROBLEM 0: What is the computational complexity of the set of avoidable words?

PROBLEM 1: How many unavoidable words are there on the alphabet with n letters? Describe these finite ordered sets.

PROBLEM 2: Is μ a recursive function?

Open Problems

PROBLEM 0: What is the computational complexity of the set of avoidable words?

PROBLEM 1: How many unavoidable words are there on the alphabet with n letters? Describe these finite ordered sets.

PROBLEM 2: Is μ a recursive function?

PROBLEM 3: Does μ achieve arbitrarily large values? Does it even get past 5?

Open Problems

PROBLEM 0: What is the computational complexity of the set of avoidable words?

PROBLEM 1: How many unavoidable words are there on the alphabet with n letters? Describe these finite ordered sets.

PROBLEM 2: Is μ a recursive function?

PROBLEM 3: Does μ achieve arbitrarily large values? Does it even get past 5?

PROBLEM 4: Let W be an avoidable word with $\mu(\text{W}) = m$. Can every W-free word on m letters be extended to a maximal W-free word on m letters?

Surveys and Monographs

*

Sapir, Mark V. *Combinatorics on Words with Applications*, Birkhäuser (To Appear).

Lothaire, M. 2002. *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, ISBN 0-521-81220-8. MR 2003i:68115

______. 1997. *Combinatorics on words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, ISBN 0-521-59924-5, With a foreword by Roger Lyndon and a preface by Dominique Perrin; Corrected reprint of the 1983 original, with a new preface by Perrin. MR 98g:68134

An Avoidable Bibliography

*

Adian, S. I. 1979. *The Burnside problem and identities in groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], vol. 95, Springer-Verlag, Berlin, ISBN 3-540-08728-1. MR 80d:20035

Allouche, Jean-Paul, James Currie, and Jeffrey Shallit. 1998. *Extremal infinite overlap-free binary words*, Electron. J. Combin. **5**, Research paper 27, 11 pp. (electronic). MR 99d:68201

Baker, Kirby A., George F. McNulty, and Walter Taylor. 1989. *Growth problems for avoidable words*, Theoret. Comput. Sci. **69**, 319–345. MR 91f:68109

Bean, Dwight R., Andrzej Ehrenfeucht, and George F. McNulty. 1979. *Avoidable patterns in strings of symbols*, Pacific J. Math. **85**, 261–294. MR 81i:20075

An Avoidable Bibliography

*

Berstel, Jean. 1979. *Sur la construction de mots sans carré*, Séminaire de Théorie des Nombres, 1978–1979, CNRS, Talence, p. Exp. No. 18, 15. MR 82a:68156 (French)

______. 1979. *Sur les mots sans carré définis par un morphisme*, Automata, languages and programming (Sixth Colloq., Graz, 1979), Lecture Notes in Comput. Sci., vol. 71, Springer, Berlin, pp. 16–25. MR 81j:68104 (French, with English summary)

______. 1980. *Mots sans carré et morphismes itérés*, Discrete Math. **29**, 235–244. MR 81e:68099 (French, with English summary)

______. 1985. *Some recent results on squarefree words*, Proceedings of the Conference on Infinitistic Mathematics (Lyon, 1984), Publ. Dép. Math. Nouvelle Sér. B, vol. 85, Univ. Claude-Bernard, Lyon, pp. 21–36. 848 821

An Avoidable Bibliography

*

Berstel, J. and P. Séébold. 1993. *A characterization of overlap-free morphisms*, Discrete Appl. Math. **46**, 275–281. MR 95d:20097 (English, with English and French summaries)

Berstel, J. 1994. *A rewriting of Fife's theorem about overlap-free words*, Results and Trends in Theoretical Computer Science (Graz, 1994), Lecture Notes in Comput. Sci., vol. 812, Springer, Berlin, pp. 19–29. MR 95c:68182

Berstel, J. and J. Karhumäki. 2003. *Combinatorics on words—a tutorial*, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS, 178–228. 1 965 433

Brandenburg, Franz-Josef. 1983. *Uniformly growing kth power-free homomorphisms*, Theoret. Comput. Sci. **23**, 69–82. MR 84i:68148

An Avoidable Bibliography

*

Brinkhuis, Jan. 1983. *Nonrepetitive sequences on three symbols*, Quart. J. Math. Oxford Ser. (2) **34**, 145–149. MR 84e:05008

Burris, S. and E. Nelson. 1971/72. *Embedding the dual of* Π_∞ *in the lattice of equational classes of semigroups*, Algebra Universalis **1**, 248–253. MR 45 #5257

Cassaigne, Julien. 1993. *Unavoidable binary patterns*, Acta Inform. **30**, 385–395. MR 94m:68155

______. 1993. *Counting overlap-free binary words*, STACS 93 (Würzburg, 1993), Lecture Notes in Comput. Sci., vol. 665, Springer, Berlin, pp. 216–225. MR 94j:68152

An Avoidable Bibliography

*

Cassaigne, Julien and James D. Currie. 1999. *Words strongly avoiding fractional powers*, European J. Combin. **20**, 725–737. MR 2000j:68154

Cassaigne, Julien. 2001. *Recurrence in infinite words (extended abstract)*, STACS 2001 (Dresden), Lecture Notes in Comput. Sci., vol. 2010, Springer, Berlin, pp. 1–11. 1 890 774

Clark, Ronald James. 2001. *Avoidable Formlas in Cominatorics on Words*, Ph.D., University of California, Los Angeles.

Currie, James D. 1984. *A direct proof of a result of Thue*, Utilitas Math. **25**, 299–302. MR 85g:05026

An Avoidable Bibliography

*

______. 1991. *Which graphs allow infinite nonrepetitive walks?*, Discrete Math. **87**, 249–260. MR 92a:05124

______. 1995. *On the structure and extendibility of k-power free words*, European J. Combin. **16**, 111–124. MR 96a:05017

______. 1996. *Non-repetitive words: ages and essences*, Combinatorica **16**, 19–40. MR 98c:68177

Currie, James D. and Robert O. Shelton. 1996. *Cantor sets and Dejean's conjecture*, J. Autom. Lang. Comb. **1**, 113–127. MR 98e:68214

An Avoidable Bibliography

*

Currie, J. and V. Linek. 2001. *Avoiding patterns in the abelian sense*, Canad. J. Math. **53**, 696–714. MR 2002g:68117

Currie, James D. and Jamie Simpson. 2002. *Non-repetitive tilings*, Electron. J. Combin. **9**, Research Paper 28, 13 pp. (electronic). MR 2003c:05053

Currie, James D. 2002. *There are ternary circular square-free words of length n for $n \geq 18$*, Electron. J. Combin. **9**, Note 10, 7 pp. (electronic). MR 2003h:05065

______. 2002. *No iterated morphism generates any Arshon sequence of odd order*, Discrete Math. **259**, 277–283. MR 2003k:05012

An Avoidable Bibliography

*

Currie, James D. and Robert O. Shelton. 2003. *The set of k-power free words over Σ is empty or perfect*, European J. Combin. **24**, 573–580. 1 983 680

Dalalyan, A. G. 1984. *Word eliminability*, Akad. Nauk Armyan. SSR Dokl. **78**, 156–158. MR 85j:05003 (Russian, with Armenian summary)

______. 1985. *X-free homomorphisms of free semigroups*, Akad. Nauk Armyan. SSR Dokl. **80**, 17–19. MR 86i:20079 (Russian, with Armenian summary)

______. 1986. *Excludable words and free homomorphisms*, Mat. Voprosy Kibernet. Vychisl. Tekhn., 76–88, 221. MR 88h:68042 (Russian, with Armenian summary)

An Avoidable Bibliography

*

Dean, Richard A. 1965. *A sequence without repeats on* x, x^{-1}, y, y^{-1}, Amer. Math. Monthly **72**, 383–385. MR 31 #3500

Dekking, F. M. 1976. *On repetitions of blocks in binary sequences*, J. Combinatorial Theory Ser. A **20**, 292–299. MR 55 #2739

Entringer, R. C., D. E. Jackson, and J. A. Schatz. 1974. *On nonrepetitive sequences*, J. Combinatorial Theory Ser. A **16**, 159–164. MR 48 #10860

Evdokimov, A. A. 1968. *Strongly asymmetric sequences generated by a finite number of symbols.*, Dokl. Akad. Nauk SSSR **179**, 1268–1271. MR 38 #3156 (Russian)

An Avoidable Bibliography

*

______. 1971. *The existence of a basis that generates 7-valued iteration-free sequences*, Diskret. Analiz, 25–30. MR 45 #4914 (Russian)

______. 1983. *Complete sets of words and their numerical characteristics*, Metody Diskret. Analiz., 7–19. MR 86e:68087 (Russian)

Hawkins, David and Walter E. Mientka. 1956. *On sequences which contain no repetitions*, Math. Student **24**, 185–187 (1957). MR 19,241a

Heitsch, Christine E. 2001. *Generalized pattern matching and the complexity of unavoidability testing*, Combinatorial Pattern Matching (Jerusalem, 2001), Lecture Notes in Comput. Sci., vol. 2089, Springer, Berlin, pp. 219–230. 1 904 579

An Avoidable Bibliography

*

______. *Intractability of the reductive decision procedure for unavoidability testing, a special case of pattern matching* (preprint).

______. *Insufficienct of four known necessary conditions on string unavoidability* (preprint).

______. *Exact distribution of deletion sizes for unavoidable strings* (preprint).

Hedlund, G. A. 1967. *Remarks on the work of Axel Thue on sequences*, Nordisk Mat. Tidskr. **15**, 148–150. MR 37 #4454

An Avoidable Bibliography

*

Istrail, Sorin. 1977. *On irreducible languages and nonrational numbers*, Bull. Math. Soc. Sci. Math. R. S. Roumanie (N.S.) **21(69)**, 301–308. MR 58 #25142

______. 1978. *Tag systems generating Thue irreductible sequences*, Information Processing Lett. **7**, 129–131. MR 58 #5110

Ježek, J. 1976. *Intervals in the lattice of varieties*, Algebra Universalis **6**, 147–158. MR 54 #7354

Karhumäki, Juhani. 1983. *On cube-free ω-words generated by binary morphisms*, Discrete Appl. Math. **5**, 279–297. MR 84j:03081

An Avoidable Bibliography

*

Larson, Jean A., Richard Laver, and George F. McNulty. 1980. *Square-free and cube-free colorings of the ordinals*, Pacific J. Math. **89**, 137–141. MR 82c:03069

Leech, Jonathan. 1957. *A problem on strings of beads*, Math. Gazette **41**, 277–278.

Lothaire, M. 1983. *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley Publishing Co., Reading, Mass., ISBN 0-201-13516-7. MR 84g:05002

______. 1997. *Combinatorics on words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, ISBN 0-521-59924-5. MR 98g:68134

An Avoidable Bibliography

*

______. 2002. *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, ISBN 0-521-81220-8. MR 2003i:68115

Margolis, Stuart W. and Mark V. Sapir. 1995. *Quasi-identities of finite semigroups and symbolic dynamics*, Israel J. Math. **92**, 317–331. MR 96i:20075

Mel′nichuk, I. L. 1984. *Undecidability of the problems of equality and divisibility in certain varieties of semigroups*, Algebra i Logika **23**, 430–438, 479. MR 86d:20065 (Russian)

______. 1984. *Solvability of problems of equality and divisibility in certain finitely based varieties of semigroups*, Properties of semigroups, Leningrad. Gos. Ped. Inst., Leningrad, pp. 91–103. MR 86j:20053 (Russian)

An Avoidable Bibliography

*

______. 1985. *Solvability of the problems of equality and divisibility in varieties of semigroups*, Theory of algebraic structures (Russian), Karagand. Gos. Univ., Karaganda, pp. 91–99. MR 89b:20122 (Russian)

Morse, Marston and Gustav A. Hedlund. 1944. *Unending chess, symbolic dynamics and a problem in semigroups*, Duke Math. J. **11**, 1–7. MR 5,202e

Novikov, P. S. and S. I. Adjan. 1968. *Infinite periodic groups. I*, Izv. Akad. Nauk SSSR Ser. Mat. **32**, 212–244. MR 39 #1532a (Russian)

______. 1968. *Infinite periodic groups. II*, Izv. Akad. Nauk SSSR Ser. Mat. **32**, 251–524. MR 39 #1532b (Russian)

An Avoidable Bibliography

*

______. 1968. *Infinite periodic groups. III*, Izv. Akad. Nauk SSSR Ser. Mat. **32**, 709–731. MR 39 #1532c (Russian)

Petrov, A. N. 1988. *A sequence that avoids every complete word*, Mat. Zametki **44**, 517–522, 558. MR 90d:68041 (Russian)

Pleasants, P. A. B. 1970. *Non-repetitive sequences*, Proc. Cambridge Philos. Soc. **68**, 267–274. MR 42 #85

Roth, Peter. 1991. *l-occurrences of avoidable patterns*, STACS 91 (Hamburg, 1991), Lecture Notes in Comput. Sci., vol. 480, Springer, Berlin, pp. 42–49. MR 92c:68112

An Avoidable Bibliography

*

______. 1992. *Every binary pattern of length six is avoidable on the two-letter alphabet*, Acta Inform. **29**, 95–107. MR 93c:68084

Sapir, M. V. 1987. *Problems of Burnside type and the finite basis property in varieties of semigroups*, Izv. Akad. Nauk SSSR Ser. Mat. **51**, 319–340, 447. MR 88h:20078 (Russian)

______. 1987. *Inherently non-finitely based finite semigroups*, Mat. Sb. (N.S.) **133(175)**, 154–166, 270. MR 88m:20116 (Russian)

Schmidt, Ursula. 1986. *Long unavoidable patterns*, STACS 86 (Orsay, 1986), Lecture Notes in Comput. Sci., vol. 210, Springer, Berlin, pp. 227–235. 827 739

An Avoidable Bibliography

*

______. 1987. *Avoidable patterns on 2 letters*, STACS 87 (Passau, 1987), Lecture Notes in Comput. Sci., vol. 247, Springer, Berlin, pp. 189–197. MR 88h:68046

______. 1987. *Long unavoidable patterns*, Acta Inform. **24**, 433–445. MR 89b:68041

______. 1989. *Avoidable patterns on two letters*, Theoret. Comput. Sci. **63**, 1–17. MR 90d:68043

Shelton, Robert. 1981. *Aperiodic words on three symbols*, J. Reine Angew. Math. **321**, 195–209. MR 82m:05004a

An Avoidable Bibliography

*

______. 1981. *Aperiodic words on three symbols. II*, J. Reine Angew. Math. **327**, 1–11. MR 82m:05004b

Shelton, Robert O. and Raj P. Soni. 1982. *Aperiodic words on three symbols. III*, J. Reine Angew. Math. **330**, 44–52. MR 82m:05004c

Shelton, R. O. 1983. *On the structure and extendibility of squarefree words*, Combinatorics on Words (Waterloo, Ont., 1982), Academic Press, Toronto, ON, pp. 101–118. MR 88m:68033

Shelton, R. O. and R. P. Soni. 1985. *Chains and fixing blocks in irreducible binary sequences*, Discrete Math. **54**, 93–99. MR 86h:68097

An Avoidable Bibliography

*

Thue, Axel. 1977. *Selected mathematical papers*, Universitetsforlaget, Oslo, ISBN 82-00-01649-8. MR 57 #46

Zech, Theodor. 1958. *Wiederholungsfreie Folgen*, Z. Angew. Math. Mech. **38**, 206–209. MR 22 #12046 (German)

Zimin, A. I. 1982. *Blocking sets of terms*, Mat. Sb. (N.S.) **119(161)**, 363–375, 447. MR 84d:20072 (Russian)