Words with the Smallest Number of Closed Factors

Gabriele Fici[1] Zsuzsanna Lipták[2]

[1] I3S, CNRS & Université Nice Sophia Antipolis, France

[2] Università di Verona, Italy

ICTCS

Varese, 21 September 2012

Closed Words

Definition

A word is closed if it is empty or it contains a factor (different from the word itself) occurring only as a prefix and as a suffix (with no internal occurrences).

Closed Words

Definition

A word is closed if it is empty or it contains a factor (different from the word itself) occurring only as a prefix and as a suffix (with no internal occurrences).

Example

The word *aabbabaab* is closed.

Closed Words

Definition

A word is closed if it is empty or it contains a factor (different from the word itself) occurring only as a prefix and as a suffix (with no internal occurrences).

Example

The word *aabbabaab* is closed.

The word *a* is closed.

Closed Words

Definition

A word is closed if it is empty or it contains a factor (different from the word itself) occurring only as a prefix and as a suffix (with no internal occurrences).

Example

The word $aabbabaab$ is closed.

The word a is closed.

A power, i.e. w^n with $n \geq 2$, is closed.

Closed Words

Definition

A word is closed if it is empty or it contains a factor (different from the word itself) occurring only as a prefix and as a suffix (with no internal occurrences).

Example

The word *aabbabaab* is closed.

The word *a* is closed.

A power, i.e. w^n with $n \geq 2$, is closed.

The words *ab* and *aabaaa* are not closed.

Closed Words

Definition

A word is closed if it is empty or it contains a factor (different from the word itself) occurring only as a prefix and as a suffix (with no internal occurrences).

Example

The word *aabbabaab* is closed.

The word *a* is closed.

A power, i.e. w^n with $n \geq 2$, is closed.

The words *ab* and *aabaaa* are not closed.

Remark

Closed words are also known as periodic-like words, or complete (first) returns.

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;
6. w has a border that does not have internal occurrences in w;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;
6. w has a border that does not have internal occurrences in w;
7. the longest border of w does not have internal occurrences in w;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;
6. w has a border that does not have internal occurrences in w;
7. the longest border of w does not have internal occurrences in w;
8. w is the complete return to its longest prefix;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;
6. w has a border that does not have internal occurrences in w;
7. the longest border of w does not have internal occurrences in w;
8. w is the complete return to its longest prefix;
9. w is the complete return to its longest border;

Closed Words

Let w be a non-empty word. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring only as a prefix and as a suffix;
2. the longest repeated prefix of w does not have internal occurrences in w;
3. the longest repeated suffix of w does not have internal occurrences in w;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;
6. w has a border that does not have internal occurrences in w;
7. the longest border of w does not have internal occurrences in w;
8. w is the complete return to its longest prefix;
9. w is the complete return to its longest border;
10. $w = uv = zu$, with v, z non-empty, and $\mathit{Fact}(w) \cap \Sigma u \Sigma = \emptyset$.

Closed Words

Exercise

Find a palindrome over two letters that is not closed.

Exercise

Find a palindrome over two letters that is not closed.

A single letter is closed...

Exercise

Find a palindrome over two letters that is not closed.

A single letter is closed...

A palindrome of length 2 is a power, so it is closed...

Closed Words

Exercise

Find a palindrome over two letters that is not closed.

A single letter is closed...

A palindrome of length 2 is a power, so it is closed...

aba is closed...

Closed Words

Exercise

Find a palindrome over two letters that is not closed.

A single letter is closed...

A palindrome of length 2 is a power, so it is closed...

aba is closed...

aabaa, *ababa*, *aabbaa* are all closed...

Closed Words

Exercise

Find a palindrome over two letters that is not closed.

A single letter is closed...

A palindrome of length 2 is a power, so it is closed...

aba is closed...

aabaa, *ababa*, *aabbaa* are all closed...

Solution: a shortest one is *aabbabaababbaa* (length 14).

Closed Factors

A factor of w that is a closed word is called a closed factor of w.

Closed Factors

A factor of w that is a closed word is called a closed factor of w.

The set of closed factors of the word w is denoted by $C(w)$.

Closed Factors

A factor of w that is a closed word is called a closed factor of w.

The set of closed factors of the word w is denoted by $C(w)$.

How many closed factors at least does a word contain?

Closed Factors

A factor of w that is a closed word is called a closed factor of w.

The set of closed factors of the word w is denoted by $C(w)$.

How many closed factors at least does a word contain?

Lemma

For any word w, one has $|C(w)| \geq |w| + 1$.

Closed Factors

A factor of w that is a closed word is called a closed factor of w.

The set of closed factors of the word w is denoted by $C(w)$.

How many closed factors at least does a word contain?

Lemma

For any word w, one has $|C(w)| \geq |w| + 1$.

Proof.

For every i, let v_i be the longest closed factor ending in position i in w. Then v_i did not occur before in w, otherwise it would not be the longest closed factor ending in position i. □

Closed Factors

A factor of w that is a closed word is called a closed factor of w.

The set of closed factors of the word w is denoted by $C(w)$.

How many closed factors at least does a word contain?

Lemma

For any word w, one has $|C(w)| \geq |w| + 1$.

Proof.

For every i, let v_i be the longest closed factor ending in position i in w. Then v_i did not occur before in w, otherwise it would not be the longest closed factor ending in position i. □

Definition

A word w is C-poor if $|C(w)| = |w| + 1$.

Closed Factors

Lemma

For any words u, v,

$$|C(u)| + |C(v)| \leq |C(uv)| + 1.$$

Lemma

For any words u, v,

$$|C(u)| + |C(v)| \leq |C(uv)| + 1.$$

We note $PAL(w)$ the set of palindromic factors of w.

Proposition (F. L., 2012)

Let w be a non-empty word of length n. If $C(w) \subseteq PAL(w)$, then $C(w) = PAL(w)$ and $|C(w)| = |PAL(w)| = n + 1$. In particular, w is a rich word.

Closed Factors

Lemma

For any words u, v,

$$|C(u)| + |C(v)| \leq |C(uv)| + 1.$$

We note $PAL(w)$ the set of palindromic factors of w.

Proposition (F. L., 2012)

Let w be a non-empty word of length n. If $C(w) \subseteq PAL(w)$, then $C(w) = PAL(w)$ and $|C(w)| = |PAL(w)| = n + 1$. In particular, w is a rich word.

Note that there are rich words that are not C-poor, e.g. *abab*.

Binary C-poor Words

We give some characterizations of binary C-poor words.

Binary C-poor Words

We give some characterizations of binary C-poor words.

Lemma

$|C(w)| = |w| + 1$ *iff every closed factor of* w *is a palindrome.*

Binary C-poor Words

We give some characterizations of binary C-poor words.

Lemma

$|C(w)| = |w| + 1$ *iff every closed factor of* w *is a palindrome.*

The lemma above is no more true on larger alphabets, e.g. *abca*.

Binary C-poor Words

We give some characterizations of binary C-poor words.

Lemma

$|C(w)| = |w| + 1$ *iff every closed factor of* w *is a palindrome.*

The lemma above is no more true on larger alphabets, e.g. *abca*.

Lemma

$|C(w)| = |w| + 1$ *iff* w *is a conjugate of a word in* $a^* b^*$.

Binary C-poor Words

So, in the binary case, we have the following characterizations:

Theorem (F. L., 2012)

Let $w \in \{a, b\}^*$. The following are equivalent:

1. $|C(w)| = |w| + 1$;
2. $C(w) = PAL(w)$;
3. $C(w) \subseteq PAL(w)$;
4. *w is a conjugate of a word in $a^* b^*$;*
5. *w does not contain any complete return to ab or ba.*

Problems

We end with the following:

Open problem

Find a characterization of C-poor words over larger alphabets.