Inner Palindromic Closure

Jürgen Dassow[1], Florin Manea[2], Robert Mercaş[2], and Mike Müller[2]

[1] Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik
[2] Christian-Albrechts-Universität zu Kiel, Institut für Informatik

Abstract. We introduce the inner palindromic closure as a new operation ♠, which consists in expanding a factor u to the left or right by v such that vu or uv, respectively, is a palindrome of minimal length. We investigate several language theoretic properties of the iterated inner palindromic closure $♠^*(w) = \bigcup_{i\geq 0} ♠^i(w)$ of a word w.

1 Introduction

The investigation of repetitions of factors in a word is a very old topic in formal language theory. For instance, already in 1906, Thue proved that there exists an infinite word over an alphabet with three letters which has no factor of the form ww. Since the eighties a lot of papers on combinatorial properties concerning repetitions of factors were published (see [4] and the references therein).

The duplication got further interest in connection with its importance in natural languages and in DNA sequences and chromosomes. Motivated by these applications, grammars with derivations consisting in "duplications" (more precisely, a word $xuwvy$ is derived to $xwuwvy$ or $xuwvwy$ under certain conditions for w, u, and v) were introduced. Combining the combinatorial, linguistic and biological aspect, it is natural to introduce the duplication language $D(w)$ associated to a word $w \in \Sigma^+$, which is the language containing all words that double some factor of w, i. e., $D(w) = \{xuuy \mid w = xuy, x, y \in \Sigma^*, u \in \Sigma^+\}$ and its iterated version $D^*(w) = \bigcup_{i\geq 0} D^i(w)$. In several papers, the regularity of $D^*(w)$ was discussed; for instance, it was shown that, for any word w over a binary alphabet, $D^*(w)$ is regular and that $D^*(abc)$ is not regular. Further cases of bounded duplication, i. e., the length of the duplicated word is bounded by a constant, were also investigated.

It was noted that words w containing hairpins, i. e., $w = xuyh(u^R)z$, and words w with $w = xuy$ and $u = h(u^R)$, where u^R is the mirror image of u and h is a letter-to-letter isomorphism, are of interest in DNA structures, where the Watson-Crick complementarity gives the isomorphism). Therefore, operations leading to words with hairpins as factors were studied (see [2]).

In this work, we consider the case where the operation leads to words which have palindromes (words with $w = w^R$) as factors (which is a restriction to the identity as the isomorphism). An easy step would be to obtain xuu^Ry from a word xuy in analogy to the duplication. But then all newly obtained palindromes are of even length. Thus it seems to be more interesting to consider the palindrome closure defined by de Luca [3]. Here a word is extended to a palindrome of minimal length. We allow this operation to be applied to factors and call it inner palindromic closure. We also study the case of iterated applications and a restriction bounding the increase of length.

For more details on basic concepts and definitions see [4].

An alphabet Σ has the cardinality denoted by $\|\Sigma\|$. A sequence of elements of Σ, called letters, constitute a word w, and we denote the empty word by ε.

For $i \geq 0$, the i-fold catenation of a word w with itself is denoted by w^i and is called the ith power of w. When $i = 2$, we call $w^2 = ww$ a square.

For a word $w \in \Sigma^*$, we denote its mirror image (or reversal) by w^R and say that w is a palindrome if $w = w^R$. For a language L, let $L^R = \{w^R \mid w \in L\}$.

2 Results

The following operation from [3] considers extensions of words into palindromes.

Definition 1. *For a word u, the left (right) palindromic closure of u is a word vu (uv) which is a palindrome for some non-empty word v such that any other palindromic word having u as proper suffix (prefix) has length greater than $|uv|$.*

As for duplication and reversal, we can now define a further operation.

Definition 2. *For a word w, the left (right) inner palindromic closure of w is the set of all words $xvuy$ ($xuvy$) for any factorisation $w = xuy$ with possibly empty x, y and non-empty u, v, such that vu (uv) is the left (right) palindromic closure of u. We denote these operations by $\spadesuit_\ell(w)$ and $\spadesuit_r(w)$, respectively, and define the inner palindromic closure $\spadesuit(w)$ as the union of $\spadesuit_\ell(w)$ and $\spadesuit_r(w)$.*

The operation is extended to languages and an iterated version is introduced.

Definition 3. *For a language L, let $\spadesuit(L) = \bigcup_{w \in L} \spadesuit(w)$. We set $\spadesuit^0(L) = L$, $\spadesuit^n(L) = \spadesuit(\spadesuit^{n-1}(L))$ for $n \geq 1$, $\spadesuit^*(L) = \bigcup_{n \geq 0} \spadesuit^n(L)$. Any set $\spadesuit^n(L)$ is called a finite inner palindromic closure of L, and we say that $\spadesuit^*(L)$ is the iterated inner palindromic closure of L.*

For the inner palindromic closure operation on binary alphabets, we get a result similar to that in [1].

Theorem 4. *The iterated inner palindromic closure of a language over a binary alphabet is regular.* □

Obviously, the finite inner palindromic closure of a finite language is always regular. However, when considering the entire class of regular languages the result is not necessarily regular.

Theorem 5. *The finite inner palindromic closure of a regular language is not necessarily regular.* □

It remains an *open problem* whether or not the iterated inner palindromic closure of a regular language L, where $\|\text{alph}(L)\| \geq 3$, is also regular.

Definition 6. *For a word u and integers $m \geq 0$ and $n > 0$, we define the sets $L_{m,n}(w) = \{u \mid u = u^R, u = xw \text{ for } x \neq \varepsilon, |x| \geq n,\ m \geq |w| - |x| \geq 0\}$, $R_{m,n}(w) = \{u \mid u = u^R, u = wx \text{ for } x \neq \varepsilon, |x| \geq n,\ m \geq |w| - |x| \geq 0\}$. The left (right) (m,n)-palindromic closure of w is the shortest word of $L_{m,n}(w)$ (resp., $R_{m,n}(w)$), or undefined if $L_{m,n}(w)$ (resp., $R_{m,n}(w)$) is empty.*

Next we investigate the (m,n)-palindromic closure, for positive integers m, n.

Theorem 7. *There exist infinitely long binary words avoiding both palindromes of length 6 and longer, and squares of words with length 3 and longer.* □

Theorem 8. *There exist infinitely long ternary words avoiding both palindromes of length 3 and longer, and squares of words with length 2 and longer.* □

We associate to a $k \geq 2$ a pair (p_k, q_k) if an infinite k-letter word avoiding palindromes of length $\geq q_k$ and squares of words of length $\geq p_k$ exists.

Theorem 9. *Let $m > 0$ and $k \geq 2$ be two integers and define $n = \max\{\frac{q_k}{2}, p_k\}$. Let Σ be a k-letter alphabet with $a \notin \Sigma$ and $w = a^m y_1 a y_2 \cdots a y_{r-1} a y_r$ be a word such that $alph(w) = \Sigma \cup \{a\}$, $r > 0$, $y_i \in \Sigma^*$ for all $1 \leq i \leq r$, and there exists j with $1 \leq j \leq r$ and $|y_j| \geq n$. Then $\spadesuit^*_{(m,n)}(w)$ is not regular.* □

The following theorem follows immediately from the previous results.

Theorem 10. *Let $w = a^p y_1 a \cdots y_{r-1} a y_r$, where $a \notin alph(y_i)$ for $1 \leq i \leq r$.*
(1) If $\|alph(w)\| \geq 3$ and $|y_j| \geq 3$ for some $1 \leq j \leq r$, then for every positive integer $m \leq p$ we have that $\spadesuit^*_{(m,3)}(w)$ is not regular.
(2) If $\|alph(w)\| \geq 4$ and $|y_j| \geq 2$ for some $1 \leq j \leq r$, then for every positive integer $m \leq p$ we have that $\spadesuit^*_{(m,2)}(w)$ is not regular.
(3) If $\|alph(w)\| \geq 5$, then for every positive integer $m \leq p$ we have that $\spadesuit^*_{(m,1)}(w)$ is not regular.
(4) For every positive integers m and n there exists u with $\spadesuit^*_{(m,n)}(u)$ not regular. □

In general, the regularity of the languages $\spadesuit^*_{(m,n)}(w)$ for positive integers m and n, and binary words w, $|w| \geq n$, is *left open*. We only show the following.

Theorem 11. *For any $w \in \{a,b\}^+$ and integer $m \geq 0$, $\spadesuit^*_{(m,1)}(w)$ is regular.* □

References

1. D. P. Bovet and S. Varricchio. On the regularity of languages on a binary alphabet generated by copying systems. *Inf. Process. Lett.*, 44:119–123, 1992.
2. D. Cheptea, C. Martín-Vide, and V. Mitrana. A new operation on words suggested by DNA biochemistry: Hairpin completion. *Trans. Comput.*, pages 216–228, 2006.
3. A. de Luca. Sturmian words: Structure, combinatorics, and their arithmetics. *Theor. Comput. Sci.*, 183:45–82, 1997.
4. G. Rozenberg and A. Salomaa. *Handbook of Formal Languages.* Springer-Verlag New York, Inc., 1997.