

CYCLIC PARTIAL WORDS

R. KRISHNA KUMARI¹, R. ARULPRAKASAM², R. PERUMAL³, AND V.R. DARE⁴

ABSTRACT. Partial words are linear words with holes. Cyclic words are derived from linear words by linking its first letter after the last one. Both partial words and cyclic words have wide applications in DNA sequencing. In this paper we introduce cyclic partial words and discuss their periodicity and certain properties. We also establish representation of a cyclic partial word using trees.

1. INTRODUCTION

DNA molecules are genetic instructions carrier. DNA strands are treated as finite strings and are used for encoding information in DNA computation. In DNA sequencing, some part of information may be absent or unseen. This can be revealed by positions denoting missing symbols in a word. Thus, instead of complete words, partial words are considered in gene comparisons. Lothaire's [9], book on combinatorics on words was the stimulus for recent works on partial words. Fischer and Paterson [5] in 1974 introduced partial words as strings with don't-care symbols which was later initiated by Berstel and Boasson [2]. F.Blanchet-Sadri [3] examined some basic combinatorial properties of words such as periodicity, conjugative property, commutative property, primitivity and showed that they also exist for partial words.

²*corresponding author*

2020 *Mathematics Subject Classification.* 68Q45.

Key words and phrases. Cyclic word, Partial word, Period.

In certain eukaryotic cells, circular or cyclic DNA sequences are found which initiated the study of cyclic words. Recently cyclic (circular) words are widely analyzed due to their usage in DNA computations similar to that of linear words. James D. Currie et al [6] introduced the study of cyclic words avoiding patterns. Laszlo Hegedus and Benedek Nagy [7, 8] defined periodicity and two ways of representations of cyclic words. Benoit Rittaud and Laurent Vivier [1] discussed certain applications of circular words. Motivated by the work on [3, 4, 7, 8, 10], the present paper extends cyclic words to cyclic partial words. The basic notions used in this paper are discussed in Section 2. In Section 3, periodicity of cyclic partial words are discussed with illustrations. In Section 4, subwords and representation of cyclic partial word using trees are discussed and finally, directions of future work is discussed in Section 5.

2. PRELIMINARIES

In this section we recall some basic definitions and notations.

Definition 2.1. Let Σ be a non-empty finite set of symbols. These symbols are termed as letters and the set is termed as an alphabet.

Definition 2.2. Any string over Σ is called a linear word. Cyclic word over Σ is derived from a linear word by linking its first letter after the last one. If u is a word of the form xy , where x is empty then $u_R = yx$ is called a rotation of u of degree $|u|$. A cyclic word is said to be primitive whenever all of its conjugates are distinct.

Definition 2.3. The sequence or word that contains a number of “do not know” symbols or “holes” denoted as \diamond are termed as partial word. The symbol \diamond is not a letter of the alphabet Σ . A partial word of length n over Σ_\diamond is a partial function $r_\diamond : \{0, 1, 2, \dots, n-1\} \rightarrow \Sigma_\diamond = \Sigma \cup \{\diamond\}$ defined by $r_\diamond(i) = r(i)$ if $i \in D(r)$, \diamond if $i \in H(r)$, where $D(r)$ and $H(r)$ are the domain set and hole set of r respectively.

Definition 2.4. A partial word r over Σ_\diamond is strongly p -periodic if there exists a non-negative integer p such that $i \equiv j \pmod{p}$ whenever $r(i) = r(j)$ for all $i, j \in D(r)$. A partial word r over Σ_\diamond is weakly p -periodic if there exists a non-negative integer p such that $r(i) = r(i+p)$ whenever $i, i+p \in D(r)$.

Definition 2.5. x is a subword of a partial word y if there exists partial words u and v such that $y = uxv$.

Definition 2.6. If x and y are two partial words of equal length and if all the elements in domain of x are also in domain of y with $x(i) = y(i)$ for all $i \in D(x)$, then x is contained in y and is denoted by $x \subset y$. Two partial words x and y are compatible, denoted by $x \uparrow y$ if $x(i) = y(i)$ for all $i \in D(x) \cap D(y)$.

3. PERIODICITY OF CYCLIC PARTIAL WORDS

In this section we define cyclic partial word and its periodicity along with few properties.

Definition 3.1. A cyclic partial word $(c_\diamond)_\diamond$ is derived from a linear partial word c_\diamond over Σ_\diamond with $|\Sigma_\diamond| \geq 2$ by linking its initial letter after the final letter. It is the set of all rotations (or conjugates) of c_\diamond . Length of $(c_\diamond)_\diamond$ denoted as $|(c_\diamond)_\diamond|$ is equal to the total number of conjugates of c_\diamond .

Example 1. Consider a partial word $u_\diamond = aa\diamond ba$ over the alphabet $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$. Then the cyclic partial word $(u_\diamond)_\diamond$ with length $|(u_\diamond)_\diamond| = 5$ is equal to

$$(aa\diamond ba)_\diamond = \{aa\diamond ba, a\diamond baa, \diamond baaa, baaa\diamond, aaa\diamond b\}.$$

Definition 3.2. The periods of a cyclic partial word are defined as follows:

- A positive integer p is a strong period of a cyclic partial word $(c_\diamond)_\diamond$ over Σ_\diamond whenever p is a period of all the conjugates of c_\diamond .
- A positive integer p is a weak period of a cyclic partial word $(c_\diamond)_\diamond$ over Σ_\diamond whenever p is a period of atleast one of the conjugates of c_\diamond .

Remark 3.1. If a cyclic partial word is strongly p periodic then it is also weakly p periodic but the converse is not true in all cases.

Theorem 3.1. If p is a strong period of a cyclic partial word, then all the multiples of p are the only strong periods of that cyclic partial word.

Proof. Let $u_\diamond \in \Sigma_\diamond^+$ be an arbitrary non-empty partial word. We have to show that if p is a smallest strong period of $(u_\diamond)_\diamond$ then $2p, 3p, \dots, np$ are the only strong periods of $(u_\diamond)_\diamond$. For instance let us consider the partial word $u_\diamond = a\diamond b\diamond ab$ over $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$. Then the cyclic partial word $(u_\diamond)_\diamond$ with length 6 is

$$(u_\diamond)_\diamond = \{a\diamond b\diamond ab, \diamond b\diamond aba, b\diamond aba\diamond, \diamond aba\diamond b, aba\diamond b\diamond, ba\diamond b\diamond a\}.$$

The periods of $(u_\diamond)_\circ$ are

$$\begin{aligned} a\diamond b\diamond ab &- 3, 4, 6 \\ \diamond b\diamond aba &- 3, 5, 6 \\ b\diamond aba\diamond &- 3, 5, 6 \\ \diamond aba\diamond b &- 3, 5, 6 \\ aba\diamond b\diamond &- 3, 5, 6 \\ ba\diamond b\diamond a &- 3, 4, 6. \end{aligned}$$

From the definition of periods of cyclic partial words, 3 and 6 are the only strong periods of $(u_\diamond)_\circ$ whereas 3, 4, 5 and 6 are weak periods. Let $p = 3$ be the minimal strong period. Then as per the statement of the theorem, 6 which is the multiple of p is the only other strong period of $(u_\diamond)_\circ$. \square

Corollary 3.1. *Let $u_\diamond \in \Sigma_\diamond^+$ be a non-primitive partial word. Then there exists strong period $p \leq |u_\diamond|/2$ of $(u_\diamond)_\circ$ and its multiple equal to $|u_\diamond|$.*

Remark 3.2. *A primitive cyclic partial word will have only one strong period equal to its length.*

Theorem 3.2. [3] *Let x, y and z be partial words such that $|x| = |y| > 0$. Then $xz \uparrow zy$ if and only if xzy is weakly $|x|$ -periodic.*

Theorem 3.3. *Let u_\diamond and v_\diamond over $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$ be two partial words of equal length. Let w_\diamond be a partial word such that $(u_\diamond w_\diamond)_\circ$ and $(w_\diamond v_\diamond)_\circ$ are weakly $|(u_\diamond)_\circ|$ -periodic whenever $(u_\diamond w_\diamond)_\circ \uparrow (w_\diamond v_\diamond)_\circ$.*

Proof. Let $u_\diamond = p_0 q_0$, $v_\diamond = q_{r+1} p_{r+2}$ and $w_\diamond = p_1 q_1 p_2 q_2, \dots, p_r q_r p_{r+1}$. Assume that $(u_\diamond w_\diamond)_\circ \uparrow (w_\diamond v_\diamond)_\circ$. Then we get

$$\begin{aligned} (u_\diamond w_\diamond)_\circ &= (p_0 q_0 p_1 q_1 p_2 q_2, \dots, p_r q_r p_{r+1})_\circ \\ &\uparrow \\ (w_\diamond v_\diamond)_\circ &= (p_1 q_1 p_2 q_2, \dots, p_{r+1} q_{r+1} p_{r+2})_\circ. \end{aligned}$$

By Theorem 3.2, $(u_\diamond w_\diamond v_\diamond)_\circ$ is also weakly $|(u_\diamond)_\circ|$ -periodic such that for all $0 \leq i \leq r+1$ and $0 \leq j \leq r+1$, $p_i \uparrow p_{i+1}$ and $q_j \uparrow q_{j+1}$. This implies that $(u_\diamond w_\diamond)_\circ$ and $(w_\diamond v_\diamond)_\circ$ are also weakly $|(u_\diamond)_\circ|$ -periodic. For instance, let $u_\diamond = a\diamond ba$, $v_\diamond =$

$\diamond a \diamond b$ such that $|u_\diamond| = |v_\diamond| = 4$. Let $w_\diamond = a \diamond b$. Then we get

$$\begin{aligned} (u_\diamond w_\diamond)_\circ &= (a \diamond baa \diamond b)_\circ \\ &\quad \uparrow \\ (w_\diamond v_\diamond)_\circ &= (a \diamond b \diamond a \diamond b)_\circ. \end{aligned}$$

The periods of $(u_\diamond w_\diamond)_\circ$ are

$$\begin{aligned} a \diamond baa \diamond b &- 4, 5, 7 \\ \diamond baa \diamond ba &- 4, 6, 7 \\ baa \diamond ba \diamond &- 4, 6, 7 \\ aa \diamond ba \diamond b &- 4, 7 \\ a \diamond ba \diamond ba &- 3, 6, 7 \\ \diamond ba \diamond baa &- 3, 6, 7 \\ ba \diamond baa \diamond &- 3, 6, 7. \end{aligned}$$

The periods of $(w^\diamond v^\diamond)_\circ$ are

$$\begin{aligned} a \diamond b \diamond a \diamond b &- 3, 4, 5, 7 \\ \diamond b \diamond a \diamond ba &- 3, 4, 6, 7 \\ b \diamond a \diamond ba \diamond &- 3, 4, 6, 7 \\ \diamond a \diamond ba \diamond b &- 3, 4, 6, 7 \\ a \diamond ba \diamond b \diamond &- 3, 4, 6, 7 \\ \diamond ba \diamond b \diamond a &- 3, 4, 6, 7 \\ ba \diamond b \diamond a \diamond &- 3, 4, 6, 7. \end{aligned}$$

we note that in $(w_\diamond v_\diamond)_\circ$, all of its conjugates have 3 as a period but it is a weak period of $(w_\diamond v_\diamond)_\circ$ because its multiple is not a period in one of the conjugates.

The periods of $(u_\diamond w_\diamond v_\diamond)_\circ$ are

$$\begin{aligned}
a \diamond baa \diamond b \diamond a \diamond b &- 4, 8, 9, 11 \\
\diamond baa \diamond b \diamond a \diamond ba &- 4, 8, 5, 10, 11 \\
baa \diamond b \diamond a \diamond ba \diamond &- 4, 8, 10, 11 \\
aa \diamond b \diamond a \diamond ba \diamond b &- 4, 8, 11 \\
a \diamond b \diamond a \diamond ba \diamond ba &- 3, 7, 10, 11 \\
\diamond b \diamond a \diamond ba \diamond baa &- 3, 6, 10, 11 \\
b \diamond a \diamond ba \diamond baa \diamond &- 3, 6, 10, 11 \\
\diamond a \diamond ba \diamond baa \diamond b &- 3, 7, 8, 10, 11 \\
\diamond a \diamond ba \diamond baa \diamond b &- 7, 8, 10, 11 \\
a \diamond ba \diamond baa \diamond b \diamond &- 7, 8, 10, 11 \\
\diamond ba \diamond baa \diamond b \diamond a &- 7, 8, 10, 11 \\
ba \diamond baa \diamond b \diamond a \diamond &- 7, 8, 10, 11.
\end{aligned}$$

$(u_\diamond w_\diamond v_\diamond)_\circ$ is weakly 3, 4, 5, 6, 7– periodic and also $(u_\diamond w_\diamond)_\circ$ and $(w_\diamond v_\diamond)_\circ$ are weakly 3, 4, 5, 6, 7– periodic which proves the theorem. \square

Property 3.1. *Two compatible cyclic partial words will have same strong periods.*

Example 2. Let $(u_\diamond)_\circ = (\diamond ba \diamond a)_\circ$ and $(v_\diamond)_\circ = (a \diamond bb \diamond)_\circ$ over $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$ with length 5 be two cyclic partial words that are compatible. The periods of u_\diamond^\diamond are

$$\begin{aligned}
\diamond ab \diamond a &- 2, 3, 4, 5 \\
ab \diamond a \diamond &- 3, 4, 5 \\
b \diamond a \diamond a &- 3, 5 \\
\diamond a \diamond ab &- 2, 4, 5 \\
a \diamond ab \diamond &- 2, 4, 5.
\end{aligned}$$

The periods of v_{\diamond}^{\diamond} are

$$\begin{aligned} a\diamond bb\diamond &- 4, 5 \\ \diamond bb\diamond a &- 2, 4, 5 \\ bb\diamond a\diamond &- 4, 5 \\ b\diamond a\diamond b &- 3, 4, 5 \\ \diamond a\diamond bb &- 4, 5. \end{aligned}$$

The above two compatible cyclic partial words have the same strong period 5.

4. SUBWORDS OF CYCLIC PARTIAL WORDS

Definition 4.1. A total word or a partial word $x \in \Sigma_{\diamond}^+$ is a subword (or factor) of a cyclic partial word $(u_{\diamond})_{\diamond}$ if x is a subword of some rotation of u_{\diamond} .

Consider a partial word c_{\diamond} of length $n \geq 3$ over $\Sigma_{\diamond} = \{a, b\} \cup \{\diamond\}$. The circular partial word $(c_{\diamond})_{\diamond}$ satisfies the following properties.

Property 4.1. For $r = 0, 1, \dots, n-2$, $(c_{\diamond})_{\diamond}$ has exactly $r+2$ subwords of length r .

Example 3. Let $c_{\diamond} = aba\diamond ab$ over $\Sigma_{\diamond} = \{a, b\} \cup \{\diamond\}$. Then

$$(c_{\diamond})_{\diamond} = \{aba\diamond ab, ba\diamond aba, a\diamond abab, \diamond ababa, ababa\diamond, baba\diamond a\}.$$

For $r = 3$, the subwords of $(c_{\diamond})_{\diamond}$ with length r are $\{aba, ba\diamond, a\diamond a, \diamond ab, bab\}$. Thus the total number of subwords is $5 = r+2$.

Property 4.2. Whenever c_{\diamond} is primitive of length $|c_{\diamond}| = n$, $(c_{\diamond})_{\diamond}$ has n subwords of length $(n-1)$ and $(n-2)$.

Example 4. Let $c_{\diamond} = \diamond ba\diamond a$ of length $|c_{\diamond}| = 5 = n$ over $\Sigma_{\diamond} = \{a, b\} \cup \{\diamond\}$. Then

$$(c_{\diamond})_{\diamond} = \{\diamond ba\diamond a, ba\diamond a\diamond, a\diamond a\diamond b, \diamond a\diamond ba, a\diamond ba\diamond\}.$$

The n number of subwords of $(c_{\diamond})_{\diamond}$ of length 3 are

$$\{\diamond ba, ba\diamond, a\diamond a, \diamond a\diamond, a\diamond b\}$$

and of length 4 are

$$\{\diamond ba\diamond, ba\diamond a, a\diamond a\diamond, \diamond a\diamond b, a\diamond ba\}.$$

Theorem 4.1. *If the tree τ of a circular partial word $(c_\diamond)_\circ$ over $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$ has one branching node on state $|(c_\diamond)_\circ| - 3$, then exists exactly one branching node on all the states $s = 0, 1, \dots, |(c_\diamond)_\circ| - 3$ of τ .*

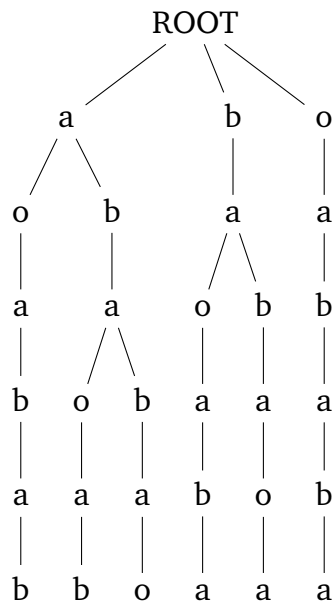
$$(aba \diamond ab)_\circ = \{aba \diamond ab, ba \diamond aba, a \diamond abab, \diamond ababa, ababa \diamond, baba \diamond a\}$$


Fig. 3: Tree of $(aba \diamond ab)_\circ$

In the above representation, there exists exactly one branching node with three edges to the parent nodes a, b and \diamond from state 0 to state 1. To prove the theorem, let us consider $b_k \in N$ branching nodes at state $k > 0$. Then, the total

number of edges in state $k + 1 =$ Total number of edges in state $k + b_k$ and the total number of edges on final state of terminal nodes $= 3 + b_1 + \dots + b_{|c_\diamond|} = |c_\diamond|$. We know that $b_i > 0$ for all $i = 1, 2, \dots, |c_\diamond| - 3$. Then $b_{|c_\diamond|} = b_{|c_\diamond|-1} = b_{|c_\diamond|-2} = 0$. Thus $3 + b_1 + \dots + b_{|w|-2} = |c_\diamond|$. Also $b_i \not\geq 1$ because for any $i > 1$, $b_j = 0$ for some $i \neq j$. Therefore $b_i = 1$ for all $i = 1, 2, \dots, |c_\diamond| - 3$. \square

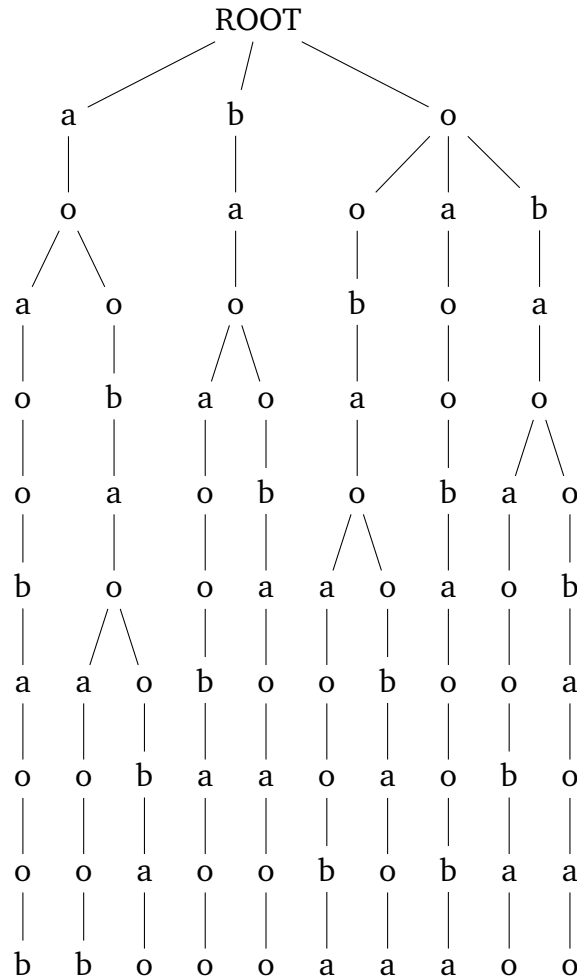
Definition 4.2. [10] Consider an alphabet Σ_\diamond with $|\Sigma_\diamond| \geq 2$. The sequence of Fibonacci partial words $\{f_n^\diamond\}$, $n \geq 0$ over the alphabet Σ_\diamond is defined as $f_{n+2}^\diamond = f_{n+1}^\diamond f_n^\diamond$. Length of Fibonacci partial word $|f_n^\diamond + 2|$ is denoted by $F^\diamond(n+2)$ such that $|f_{n+2}^\diamond| = |f_{n+1}^\diamond| + |f_n^\diamond|$ for all $n \geq 0$ and $f_0^\diamond, f_1^\diamond$ are initial Fibonacci partial words with $|f_0^\diamond|, |f_1^\diamond| \geq 2$.

Example 5. Consider $f_0^\diamond = \diamond b$ and $f_1^\diamond = aa\diamond$ as initial Fibonacci partial words where $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$. Here $|f_0^\diamond| = 2$ and $|f_1^\diamond| = 3$. Then the sequence of Fibonacci partial words $\{f_5^\diamond\}$ are as follows:

$$\begin{aligned} f_0^\diamond &= \diamond b \\ f_1^\diamond &= aa\diamond \\ f_2^\diamond &= aa\diamond\diamond b \\ f_3^\diamond &= aa\diamond\diamond baa\diamond \\ f_4^\diamond &= aa\diamond\diamond baa\diamond aa\diamond\diamond b \\ f_5^\diamond &= aa\diamond\diamond baa\diamond aa\diamond\diamond baa\diamond\diamond baa\diamond. \end{aligned}$$

Corollary 4.1. Let ξ_n represent the tree of the circular Fibonacci partial word $(f_n^\diamond)_\circ$ where the sequence $\{f_n^\diamond\}$ has equal initial lengths say $|f_0^\diamond| = |f_1^\diamond|$. Then there exists exactly one branching node on all the states $s = 0, 1, \dots, |f_n^\diamond| - 2|f_0^\diamond|$ of ξ_n .

Example 6. Consider $f_0^\diamond = \diamond b$ and $f_1^\diamond = a\diamond$ as initial Fibonacci partial words where $\Sigma_\diamond = \{a, b\} \cup \{\diamond\}$. Here $|f_0^\diamond| = 2$ and $|f_1^\diamond| = 2$. Then $(f_4^\diamond)_\circ = (a\diamond\diamond ba\diamond a\diamond\diamond b, \diamond\diamond ba\diamond a\diamond\diamond ba, \diamond ba\diamond a\diamond\diamond ba\diamond, ba\diamond a\diamond\diamond ba\diamond\diamond, a\diamond a\diamond\diamond ba\diamond\diamond b, \diamond a\diamond\diamond ba\diamond\diamond ba, a\diamond\diamond ba\diamond\diamond ba\diamond, \diamond\diamond ba\diamond\diamond ba\diamond a, \diamond ba\diamond\diamond ba\diamond a\diamond, ba\diamond\diamond ba\diamond a\diamond\diamond)$.



In the above tree, there exists exactly one branching node on all the states $s = 0, 1, \dots, 6$

Example 7. Consider Example 6. Here the tree of $(f_3^\diamond)_\circ$ also appears in the tree of $(f_4^\diamond)_\circ$.

5. CONCLUSION

In this paper we introduced cyclic partial words and discussed their periodicity and certain properties. We also established representation of a cyclic partial

word using trees. In future, other combinatorial properties of partial word can be extended to cyclic partial words. Also Tree representation can be used for the study of subwords of cyclic partial words.

REFERENCES

- [1] B. RITTAUD, L. VIVIER: *Circular words and applications*, Proceedings of Words 2011, Electronic Proceedings in Theoretical Computer Science, **63** (2011), 31-36.
- [2] J. BERSTEL, L. BOASSON: *Partial Words and a Theorem of Fine and Wilf*, Theoret. Comput.Sci., **218** (1999), 135-141.
- [3] F. BLANCHET-SADRI: *Algorithmic Combinatorics on Partial Words*, Discrete Mathematics and its Applications, CRC Press, 2007.
- [4] D. NOWOTKA: *Periodicity and Unbordered Factors of Words*, TUCS Dissertations No.50, 2004.
- [5] M.J. FISCHER, M.S. PATERSON: *String Matching and other Products*, In: Karp, R.M. (ed.) Complexity of Computation. SIAM-AMS Proceedings, (1974), 113-125.
- [6] J.D. CCURIE, D.S. FITZPATRICK: *Circular Words Avoiding Patterns*, Proceedings of the 6th International Conference on Developments in Language Theory. LNCS, 2450 (2002), 319-325.
- [7] L. HEGEDUS, B. NAGY: *Representations of Circular Words*, In: Automata and Formal Languages, (2014), 261-270.
- [8] L. HEGEDUS, B. NAGY: *Periodicity of Circular Words*, Local Proceedings of WORDS TUCS Lecture Notes, **20** (2013), 45-56.
- [9] M. LOTHAIRE: *Combinatorics on Words*, Addison-Wesley, 1983.
- [10] R.K. KUMARI, R. ARULPRAKASAM, V.R. DARE: *Combinatorial Properties of Fibonacci Partial Words and Arrays*, Journal of Discrete Mathematical Sciences and Cryptography, (Accepted) (2020).

^{1,2,3}DEPARTMENT OF MATHEMATICS, SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, KATTANKULATHUR-603 203, TAMILNADU, INDIA.

Email address: ¹kr1062@srmist.edu.in

Email address: ²r.aruljeeva@gmail.com

Email address: ³perumalnew@yahoo.co.in

⁴DEPARTMENT OF MATHEMATICS, MADRAS CHRISTIAN COLLEGE, CHENNAI-600 059, TAMILNADU, INDIA.

Email address: ⁴rajkumardare@yahoo.com