m-апериодические слова над трехбуквенным алфавитом

В. И. Сенашов

Институт вычислительного моделирования СО РАН
Российская Федерация, 660036, Красноярск, Академгородок, 50/44
E-mail: sen1112home@mail.ru

Работа посвящена изучению множеств апериодических слов над конечным алфавитом. Множество апериодических слов можно рассматривать как словарь некоторого конечного формального языка. Существование бесконечных слов в двухбуквенном или трехбуквенном алфавитах, которые не содержат подслов, являющихся третьими степенями или, соответственно, квадратами других слов, впервые получены более ста лет назад. С. И. Адян в 2010 г. построил пример бесконечной последовательности несократимых слов, каждое из которых является началом следующего и не содержит квадратов слов в алфавите из двух букв. С. Е. Аршон установил существование n-значной ассиметричной бесповторной последовательности для алфавита не менее чем из трех букв. В монографии С. И. Адяна доказано, что в алфавите из двух символов существуют бесконечные 3-апериодические последовательности. В работах других авторов рассматривались обобщения апериодичности, когда исключались не только степени некоторых подслов. В монографии А. Ю. Ольшанского доказана бесконечность множества 6-апериодических слов в двухбуквенном алфавите и получена оценка количества таких слов любой данной длины. Автором ранее случай трехбуквенного алфавита рассмотрен только в случае 6-апериодических слов. В данной статье доказана бесконечность множества m-апериодических слов в трехбуквенном алфавите при $m \geq 4$ и получена оценка множества таких слов. Полученные результаты могут быть полезны при кодировании информации в сеансах космической связи.

Ключевые слова: алфавит, бесповторная последовательность, слово, апериодичность, оценка, формальный язык.

m-aperiodic words on three-letter alphabet

V. I. Senashov

Institute of Computational Modelling of Siberian Branch of RAS
50/44, Akademgorodok, Krasnoyarsk, 660036, Russian Federation
E-mail: sen1112home@mail.ru

The work is devoted to the study of sets of aperiodic words over a finite alphabet. The set of aperiodic words can be considered as a dictionary of some finite formal language. The existence of infinite words in two-letter or three-letter alphabets that do not contain subwords that are third powers or, respectively, squares of other words was first discovered more than a hundred years ago. S.I. Adyan in 2010 constructed an example of an infinite sequence of irreducible words, each of which is the beginning of the next and does not contain word squares in a two-letter alphabet. S.E. Arshon established the existence of an n-digit

asymmetric repetition-free sequence for an alphabet of at least three letters. In the monograph by S.I. Ad-yan proved that in an alphabet of two symbols there exist infinite 3-aperiodic sequences. In the works of other authors, generalizations of aperiodicity were considered, when not only the powers of some subwords were excluded. In the monograph by A.Yu. Olshansky proved the infinity of the set of 6-aperiodic words in a two-letter alphabet and obtained an estimate for the number of such words of any given length. The author previously considered the case of a three-letter alphabet only in the case of 6-aperiodic words. In this article, we prove the infinity of the set of m-aperiodic words in the three-letter alphabet at $m \geq 4$ and obtain an estimate for the set of such words. The results can be applied when encoding information in space communications.

Keywords: alphabet, non-repetitive sequence, word, aperiodicity, estimate, formal language.

Introduction

The work is devoted to the study of sets of aperiodic words over a finite alphabet. The set of aperiodic words can be considered as a dictionary of some finite formal language.

Data on the existence of an infinite word in a two- or three-letter alphabet, which does not contain subwords that are cubes or, respectively, squares, were first obtained by A. Thue in 1906. [1]. An example of an infinite sequence of irreducible words, each of which is the beginning of the next and does not contain squares of words in an alphabet of two letters, was constructed by S. I. Adyan (Lemma 1 from [2]). In an article by S. E. Arshon in 1937. [3] the existence of an n-digit asymmetric repetition-free sequence for an alphabet of at least three letters has been proven. In the monograph by S. I. Adyan [4] it has been proven that in an alphabet of two symbols there exist infinite 3-aperiodic sequences. The problem of studying such sequences is considered by S. I. Adyan in connection with his study of the Burnside problem [5–7]. A. M. Shur found out which properties of ordinary words are preserved during the transition to infinite ones, and which ones are modified, lost or replaced by new ones, and studied the set of all cubeless Z-words in a two-letter alphabet [8–10]. In the monograph by A. Yu. Olshansky [11], the infinity of the set of 6-aperiodic words in a two-letter alphabet was proved and an estimate was obtained for the number of such words of any given length.

We have improved A. Yu. Olshansky's estimate [11] of the number of 6-aperiodic words in the 2-letter alphabet [12], and also made a report on the topic of aperiodic words [13]. Then research on this issue continued.

In [14], a theorem was proven about the infinity of the set of m-aperiodic words for $m \geq 4$ in a two-letter alphabet and an estimate of the number of such words was obtained. The case of a three-letter alphabet was considered only in one case: in [15], an estimate was obtained for the number of 6-aperiodic words.

In this article, the set of m-aperiodic words in the three-letter alphabet will be studied: the infinity of the set of *m*-aperiodic words for $m \geq 4$ *in* three letters alphabet and an estimate of the number of such words was obtained. The results may be useful when encoding information in space communication sessions.

Generalizations of aperiodicity, when not only the powers of some subwords are excluded, were studied in [16].

Main result

We consider that a *periodic word with period* H is any subword of some word H^p, $p > 0$.

As an example of a periodic word with period *ab,* we consider the word *ababa* .

A *l-aperiodic word* is a word X in which there are no nontrivial subwords of the form Y^l.

S.I. Adyan [4] proved that in an alphabet of two letters there are infinitely many arbitrarily long 3-aperiodic words.

S. E. Arshon established the existence of words of any length free from squares in a three-letter alphabet [3].

A. Yu. Olshansky in [11] considered a set of 6-aperiodic words. He proved that such words of any length exist and obtained an estimate for the function $f(n)$ – number of 6-aperiodic words of length n. There were more of them than $\left(\frac{3}{2}\right)^n$. In [12], we improved A. Yu. Olshansky's estimate of the number of 6-aperiodic words over a 2-letter alphabet.

In [15], we proved a theorem about the infinity of the set of m-aperiodical words under a period limitation $m \geq 4$ in a two-letter alphabet and obtained a lower estimate for the number of such words.

It is of interest to us to estimate the number of *m*-aperiodic words in an alphabet of three letters.

When proving the theorem, we will use the method used by A. Yu. Olshansky [11].

Theorem. *In the three-letter alphabet there are arbitrarily long m-aperiodical words for $m \geq 4$ and number $f(n)$ there are more such words of length n than $\left(\frac{5}{2}\right)^n$.*

Proof. We consider the three-letter alphabet {*a, b, c*}. In this alphabet we will study m-aperiodic words for $m \geq 4$. We will investigate the function $f(n)$ number of *m*-aperiodic words of length *n*.

First we prove the inequality $f(n+1) > (5/2)f(n)$ using the method of mathematical induction.

We note that there are exactly three *m*-aperiodic words *a*, *b*, *c* of length one and nine *m*-aperiodic words *aa, ab, ac, ba, bb, bc, ca, cb, cc* of length 2 in the three-letter alphabet for $m \geq 4$. Thus, in this case we have $f(1) = 3$, $f(2) = 9$.

Then for *n* = 1 the inequality holds $f(2) > \frac{5}{2} \cdot f(1)$ and the resulting estimate is as a base of induction

Now we need to take the induction step. We assume that the inequality $f(n) > (5/2)f(n-1)$ is true for all values not exceeding *n*, and establish its validity in the case $f(n+1) > (5/2)f(n)$.

Any *m*-aperiodic word of length $n+1$ can be obtained by assigning the letters *a*, *b* or *c* to the right of the *m*-aperiodic word of lengt *n*. This produces $3f(n)$ words of *X* length $n+1$.

Some of these words contain degrees A^m. We try to discard such words containing subwords of period *m*.

When assigning a new letter to the right, you can only get a word like $X \equiv YA^m$, containing a word of period *m*, since otherwise the beginning of length *n* of the word *X* of length $n+1$ contains a periodic subword of period m: A^m.

For words *A* of length 1 (there are only three such words), there are at most $3f(n-m+1)$ words of $X \equiv YA^m$ type, where the word *Y* *m is* aperiodic and $|Y| = n-m+1$, *A* is one of the words *a*, *b*, *c*.

As shown above, there are only nine words *A* of length 2. Then the number of words of the form $X \equiv YA^m$ length $n+1$ is no more than $9f(n-2m+1)$, where the word *Y is* *m*-aperiodic and has a length $n-2m+1$.

We further reason in the same way and obtain an estimate for the number of words of length $n + 1$ (a strict inequality, since it is easy to indicate words that, when obtaining an estimate for the number of words of the form $X \equiv YA^m$, already with a length of 1 word A, we discarded in excess to obtain an estimate):

$$f(n+1) > 3f(n) - 3f(n-m+1) - 3^2 f(n-2m+1) - 3^3 f(n-3m+1) - \ldots$$

Since the estimate is valid by inductive assumption $f(n) > \left(\frac{5}{2}\right)^k \cdot f(n-k)$, applying it, we see:

$$f(n+1) > 3f(n) - \left(3\left(\frac{5}{2}\right)^{-m+1} f(n) + 3^2 \left(\frac{5}{2}\right)^{-2m+1} f(n) + 3^3 \left(\frac{5}{2}\right)^{-3m+1} f(n) + \ldots\right).$$

We transform the resulting inequality:

$$f(n+1) > f(n)(3 - \left(3\left(\tfrac{5}{2}\right)^{-m+1} + 3^2\left(\tfrac{5}{2}\right)^{-2m+1} + 3^3\left(\tfrac{5}{2}\right)^{-3m+1} + ... \right).$$

Obviously, $\sqrt[m]{3} < \tfrac{5}{2}$, therefore the geometric progression in brackets is decreasing with the denominator $3\left(\tfrac{5}{2}\right)^{-m}$ with the limitation $m \geq 4$.

For the expression S in parentheses on the right side, we apply the formula for the sum of a decreasing geometric progression

$$S = 3 - \frac{3\left(\tfrac{5}{2}\right)^{-m+1}}{1 - 3\left(\tfrac{5}{2}\right)^{-m}}.$$

To prove the theorem, we need an estimate of the expression in brackets on the right side of the inequality $S > \tfrac{5}{2}$.

We substitute the expression obtained above for S into the inequality and perform elementary transformations:

$$\frac{3 - 9\left(\tfrac{5}{2}\right)^{-m} - 3\left(\tfrac{5}{2}\right)^{-m+1} + 3\left(\tfrac{5}{2}\right)^{-m+1} - \tfrac{5}{2}}{1 - 3\left(\tfrac{5}{2}\right)^{-m}} > 0.$$

We will simplify the resulting inequality:

$$\frac{3 - 9\left(\tfrac{5}{2}\right)^{-m} - \tfrac{5}{2}}{1 - 3\left(\tfrac{5}{2}\right)^{-m}} > 0.$$

Since the inequality $1 - 3\left(\tfrac{5}{2}\right)^{-m} > 0$ holds for $m \geq 4$, it remains to check the validity of the inequality :

$$3 - 9\left(\tfrac{5}{2}\right)^{-m} - \tfrac{5}{2} > 0.$$

The validity of this inequality is established by direct calculation at $m = 4$, and $m \geq 4$ at it is even more true, since $\left(\tfrac{5}{2}\right)^{-m} < \left(\tfrac{5}{2}\right)^{-3}$ for such *m*. Consequently, it has been proven that the inequality $f(n+1) > \tfrac{5}{2} f(n)$ holds for any natural values of the number *n*, subject to $m \geq 4$.

The theorem is proven.

Conclusion

The question of estimating the number of aperiodic words under various conditions continues to be studied. The set of *m*-aperiodic words for $m \geq 4$ in a three-letter alphabet is considered and an estimate is obtained for the function of the number of such words of any given length.

Библиографические ссылки

1. Thue A. Uber unendliche Zeichenreih // Norcke Vid. Selsk. skr., I Mat. Nat. Kl. Christiania. 1906. Bd. 7. P. 1–22.

2. Адян С. И. Проблема Бернсайда и связанные с ней вопросы // Успехи мат. наук. 2010. Т. 65, вып. 5 (395). С. 5–60.

3. Аршон С. Е. Доказательство существования *n*-значных бесконечных асимметричных последовательностей // Мат. сб. 1937. № 4 (2 (44)). С. 769–779.

4. Адян С. И. Проблема Бернсайда и тождества в группах. М. : Наука. 1975. 336 с.

5. Новиков П. С., Адян С. И. О бесконечных периодических группах. I // Изв. АН СССР. Сер. мат. 1967. № 1 (32). С. 212–244.

6. Новиков П. С., Адян С. И. О бесконечных периодических группах II // Изв. АН СССР. Сер. мат. 1967. № 2 (32). С. 251–524.

7. Новиков П. С., Адян С. И. О бесконечных периодических группах III // Изв. АН СССР. Сер. мат. 1967. № 3 (32). С. 708–731.

8. Шур А. М. Структура множества бескубных *Z*-слов в двухбуквенном алфавите // Изв. РАН. Сер. мат. 2000. Вып. 4 (64). С. 201–224.

9. Shur A. M. Overlap-free words and Thue-Morse sequences // Int. J. Alg. and et al. 1996. Vol. 6. P. 353–367.

10. Shur A. M. Binary words avoided by the Thue-Morse sequence // Semigroup Forum. 1996. Vol. 53. P. 212–219.

11. Ольшанский А. Ю. Геометрия определяющих соотношений в группах. М. : Наука. 1989. 448 с.

12. Сенашов В. И. Улучшение оценки количества 6-апериодических слов фиксированной длины // Вестник СибГАУ. 2016. № 2 (17). С. 168–172.

13. Сенашов В. И. Апериодические слова // Решетневские чтения. 2015. Т. 2, № 19. С. 132–133.

14. Senashov V. I. Estimation of the number of aperiodic words // Сибирский аэрокосмический журнал. 2022. № 3 (23). С. 409–416 .

15. Senashov V. I. 6-aperiodic words over the three-letter alphabet // Сибирский журнал науки и технологий. 2020. № 3 (21). С. 333–336.

16. Bean D. R., Ehrenfeucht A., McNulty G. F. Avoidable patterns in strings of symbols // Pacific J. Math. 1979. No. 2 (85). P. 261–295.

References

1. Thue A. Uber unendliche Zeichenreih. Norcke Vid. Selsk. skr., I Mat. Nat. Kl. Christiania. 1906, bd. 7, p. 1–22.

2. Adyan S. I. [Burnside's problem and related questions]. *Uspekhi Mat. sciences*. 2010, Vol. 65, Is. 5 (395), P. 5–60. (In Russ.)

3. Arshon S. E. [Proof of existence of *n*-unit infinite asymmetric sequences]. *Mat. sb*. 1937, No. 4 (2 (44)), P. 769–779. (In Russ.)

4. Adyan S. I. *Problema Bernsayda i tozhdestva v gruppakh* [Bernside Problem and Identities in Groups]. Moscow, Nauka Publ., 1975, 336 p.

5. Novikov P. S., Adyan S. I. [On infinite periodic groups. I]. *Izv. USSR Academy of Sciences, ser. math*. 1967, No. 1 (32), P. 212–244. (In Russ.)

6. Novikov P. S., Adyan S. I. [On infinite periodic groups II]. *Izv. USSR Academy of Sciences, ser. math.* 1967, No. 2 (32), P. 251–524. (In Russ.)

7. Novikov P. S., Adyan S. I. [On infinite periodic groups III]. *Izv. USSR Academy of Sciences, ser. math*. 1967, No. 3 (32), P. 708–731. (In Russ.)

8. Shur A. M. [Structure of the set of cubeless Z-words in a two-letter alphabet]. *Izv. RAS. Ser. Mat*. 2000, Vol. 64, Is. 4, P. 201–224. (In Russ.)

9. Shur A. M. Overlap-free words and Thue-Morse sequences. *Int. J. Alg. and et al.* 1996. Vol. 6. P. 353–367.

10. Shur A. M. Binary words avoided by the Thue-Morse sequence. *Semigroup Forum*. 1996. Vol. 53. P. 212–219.

11. Olshansky A. Yu. *Geometriya opredelyayushchikh sootnosheniy v gruppakh* [Geometry of defining relations in groups]. Moscow, Nauka Publ., 1989, 448 p.

12. Senashov V. I. [Improved estimates of the number 6-aperiodic words of fixed length]. *Vestnik SibGAU*. 2016, No. 2 (17), P. 168–172 (In Russ.).

13. Senashov V. I. [Aperiodic words]. *Reshetnevskiye chteniya*. 2015, Vol. 2, No. 19, P. 132–133. (In Russ.)

14. Senashov V. I. [Estimation of the number of aperiodic words]. *Siberian Aerospace Journal.* 2022. No. 3 (23), P. 409–416. (In Russ.)

15. Senashov V. I. [6-aperiodic words over the three-letter alphabet]. *Siberian Journal of Science and Technology*. 2020, No. 3 (21), P. 333–336. (In Russ.)

16. Bean D. R., Ehrenfeucht A., McNulty G. F. Avoidable patterns in strings of symbols. *Pacific J. Math.* 1979, No. 2 (85), P. 261–295.

Сенашов Владимир Иванович – доктор физико-математических наук, профессор, ведущий научный сотрудник; Институт вычислительного моделирования СО РАН. E-mail: sen1112home@mail.ru.

Senashov Vladimir Ivanovich – doctor of physic and mathematic sciences, professor, leader researcher; Institute of Computational Modelling of Siberian Branch of RAS. E-mail: sen1112home@mail.ru.