Journal of Integer Sequences, Vol. 16 (2013),
Article 13.2.1

Clustering Words and Interval Exchanges

Sébastien Ferenczi
IMPA
CNRS — UMI 2924
Estrada Dona Castorina 110
Rio de Janeiro, RJ 22460-32
Brazil
ferenczi_sebastien@yahoo.fr

Luca Q. Zamboni
Institut Camille Jordan
Université Claude Bernard Lyon 1
43 boulevard du 11 novembre 1918
F-69622 Villeurbanne Cedex
France
and
Department of Mathematics
and Turku Centre for Computer Science
University of Turku
20014 Turku
Finland
zamboni@math.univ-lyon1.fr

Abstract

We characterize words which cluster under the Burrows-Wheeler transform as those words w such that ww occurs in a trajectory of an interval exchange transformation, and build examples of clustering words.

1 Introduction

In 1994 Michael Burrows and David Wheeler [1] introduced a transformation on words which proved very powerful in data compression. The aim of the present note is to characterize those words which cluster under the Burrows-Wheeler transform, that is to say, those words that are transformed into expressions such as $4^a3^b2^c1^d$ or $2^a5^b3^c1^d4^e$. Clustering words on a binary alphabet have already been extensively studied (see, for instance, [8, 11]) and identified as particular factors of the Sturmian words. Some generalizations and partial characterizations to r letters appear in Restivo and Rosone [14], but it had not yet been observed that clustering words are intrinsically related to a dynamical object called *interval exchange transformations* introduced in Oseledets [12]: we shall define them in Definitions 1 and 2 below, and refer the reader to [16] which constitutes a classical course on general interval exchange transformations and contains many of the technical terms found in Section 4 below. This link comes essentially from the fact that the array of conjugates used to define the Burrows-Wheeler transform gives rise to a *discrete* interval exchange transformation sending its first column to its last column. It turns out that the converse is also true: interval exchange transformations generate clustering words. Indeed we prove that clustering words are exactly those words w such that ww occurs in a trajectory of an interval exchange transformation. On a binary letter alphabet, this condition amounts to saying that ww is a factor of an infinite Sturmian word. We end the paper by some examples and questions on how to generate clustering words.

This paper began during a workshop on board Via Rail Canada train number 2. We are grateful to Laboratoire International Franco-Québécois de Recherche en Combinatoire (LIRCO) for funding and Via for providing optimal working conditions. The second author is partially supported by a FiDiPro grant from the Academy of Finland.

The authors owe much to Jean-Paul Allouche, the first mathematician who revealed to both of them the beauty of combinatorics on words, and who taught the first author that not every paper needs to begin with "Let (X, T, μ) be a system ...".

2 Definitions

Let $A = \{a_1 < a_2 < \cdots < a_r\}$ be an ordered alphabet and $w = w_1 \cdots w_n$ a *primitive* word on the alphabet A, i.e., w is not a power of another word. For simplification we suppose that *each letter of* A *occurs in* w.

The *Parikh vector* of w is the integer vector $(n_1, \dots, n_k)$ where n_i is the number of occurrences of a_i in w. The *(cyclic) conjugates* of w are the words $w_i \cdots w_n w_1 \cdots w_{i-1}$, $1 \leq i \leq n$. As w is primitive, w has precisely n-cyclic conjugates. Let $w_{i,1} \cdots w_{i,n}$ denote the i-th conjugate of w where the n-conjugates of w are ordered by ascending lexicographical order. Then the *Burrows-Wheeler transform* of w, denoted by $B(w)$, is the word $w_{1,n}w_{2,n} \cdots w_{n,n}$. In other words, $B(w)$ is obtained from w by first ordering its cyclic conjugates in ascending order in a rectangular array, and then reading off the last column. For instance $B(2314132) = 4332211$. We say w is π-*clustering* if $B(w) = a_{\pi 1}^{n_{\pi 1}} \cdots a_{\pi r}^{n_{\pi r}}$, where $\pi \neq Id$ is a permutation on $\{1, \dots, r\}$. We say w is *perfectly* clustering if it is π-clustering for $\pi i = r + 1 - i$, $1 \leq i \leq r$. For instance 2314132 is perfectly clustering. Restivo and Rosone [14] showed that if w perfectly clusters, then w is strongly (or circularly) rich, i.e., w^2 has $|w^2|+1$ distinct palindromic factors. But this condition is not a characterization of perfectly clustering words (see Example 6.4 in Restivo and Rosone[14]).

Definition 1. A (continuous) *r-interval exchange transformation* T with probability vector $(\alpha_1, \alpha_2, \ldots, \alpha_r)$, and permutation π is defined on the interval $[0,1[$, partitioned into r intervals

$$\Delta_i = \left[\sum_{j<i} \alpha_j, \sum_{j\le i} \alpha_j\right[,$$

by

$$Tx = x + \tau_i \quad \text{when} \quad x \in \Delta_i,$$

where $\tau_i = \sum_{\pi^{-1}(j)<\pi^{-1}(i)} \alpha_j - \sum_{j<i} \alpha_j$.

Intuitively this means that the intervals Δ_i are re-ordered by T following the permutation π. Note that our use of the word "continuous" does not imply that T is a continuous map on $[0,1[$ (though it can be modified to be made so); it is there to emphasize the difference with its discrete analogue.

Definition 2. A *discrete r-interval exchange transformation* T with length vector

$$(n_1, n_2, \ldots, n_r),$$

and permutation π is defined on a set of $n_1 + \cdots + n_r$ points $x_1, \ldots, x_{n_1+\cdots+n_r}$ partitioned into r intervals

$$\Delta_i = \{x_k, \sum_{j<i} n_j < k \le \sum_{j\le i} n_j\}$$

by

$$Tx_k = x_{k+s_i} \quad \text{when} \quad x_k \in \Delta_i,$$

where $s_i = \sum_{\pi^{-1}(j)<\pi^{-1}(i)} n_j - \sum_{j<i} n_j$.

We recall the following notions, defined for any transformation T on a set X equipped with a partition Δ_i, $1 \le i \le r$.

Definition 3. The *trajectory* of a point x under T is the infinite sequence $(x_n)_{n\in\mathbb{N}}$ defined by $x_n = i$ if $T^n x$ belongs to Δ_i, $1 \le i \le r$. The mapping T is *minimal* if whenever E is a nonempty closed subset of X and $T^{-1}E = E$, then $E = X$.

3 Main result

Theorem 4. *Let $w = w_1 \cdots w_n$ be a primitive word on $A = \{1, \ldots, r\}$, such that every letter of A occurs in w. The following are equivalent:*

1. *w is π-clustering,*

2. *ww occurs in a trajectory of a minimal discrete r-interval exchange transformation with permutation π,*

3. *ww occurs in a trajectory of a discrete r-interval exchange transformation with permutation π,*

4. *ww occurs in a trajectory of a continuous r-interval exchange transformation with permutation π.*

Proof. ((2), (3) or (4) implies (1)) By assumption there exists a point x whose initial trajectory of length $2n$ is the word ww. Consider the set $E = \{Tx, T^2x, \dots, T^nx\}$. Then for each $y \in E$, the initial trajectory of y of length n, denoted $O(y)$, is a cyclic conjugate of w.

Suppose y and z are in E, and y is to the left of z (meaning $y < z$.). Let j be the smallest nonnegative integer such that T^jy and T^jz are not in the same Δ_i. Then T^jy is to the left of T^jz, either because $j = 0$ or because T is increasing on each Δ_i. Thus $O(y)$ is lexicographically smaller that $O(z)$.

Thus $B(w)$ is obtained from the last letter $l(y)$ of $O(y)$ where the points y are ordered from left to right. But $l(y)$ is the label of the interval Δ_i where $T^{n-1}y$, or equivalently $T^{-1}y$, falls. Thus by definition of T, if y is to the left of z then $\pi^{-1}(l(y)) \leq \pi^{-1}(l(z))$, and if y' is between y and z with $l(y) = l(z)$, then $l(y') = l(y) = l(z)$, hence the claimed result. □

Proof. ((2) implies (3) implies (4)) The first implication is trivial. The second follows from the fact that the trajectories of the discrete r-interval exchange transformation with length vector $(n_1, n_2, \dots, n_r)$, and permutation π, and of the continous r-interval exchange transformation with probability vector $(\frac{n_1}{n_1+\dots+n_r}, \dots, \frac{n_r}{n_1+\dots+n_r})$ and permutation π are the same. We note that this continuous interval exchange transformation is never minimal, while the discrete one may be. □

We now turn to the proof of the converse, which uses a succession of lemmas. Throughout this proof, unless otherwise stated, a given word w is a primitive word on $\{1, \dots, r\}$, and every letter of $\{1, \dots, r\}$ occurs in w; $(n_1, \dots, n_r)$ is its Parikh vector, the $w_{i,1} \cdots w_{i,n}$ are its conjugates.

The first lemma states that B is injective on the conjugacy classes, which is proved for example in Crochemore, Désarménien, and Perrin [2] or Mantaci, Restivo, Rosone and Sciortino citerr1; we give here a short proof for sake of completeness.

Lemma 5. *If w and w' are words such that $B(w) = B(w')$, then w and w' are cyclically conjugate*

Proof. In the array of the conjugates of w, each column word $w_{1,j} \cdots w_{n,j}$ has the same Parikh vector as w, so we retrieve this vector from $B(w)$; thus we know the first column word, which is $1^{n_1} \dots r^{n_r}$, and the last column word which is $B(w)$. Then the words $w_{n,j}w_{1,j}$ are precisely all words of length 2 occurring in the conjugates of w, and by ordering them we get the first two columns of the array. Then $w_{n,j}w_{1,j}w_{2,j}$ constitute all words of length 3 occurring in the conjugates of w, and we get also the subsequent column, and so on until we have retrieved the whole array, thus w up to conjugacy. □

It is easy to see that B, viewed as a mapping from words to words, is not surjective (see for instance [10]). A more precise result will be proved in Corollary 7 below.

Lemma 6. *If w is π-clustering, the mapping $w_{1,j} \mapsto w_{n,j}$ defines a discrete r-interval exchange transformation with length vector $(n_1, n_2, \ldots, n_r)$, and permutation π.*

Proof. We order the occurrences of each letter in w by putting $w_i < w_j$ if the conjugate $w_i \cdots w_n w_1 \cdots w_{i-1}$ is lexicographically smaller than $w_j \cdots w_n w_1 \cdots w_{j-1}$. By primitivity, the n letters of w are uniquely ordered as

$$1_1 < \cdots < 1_{n_1} < 2_1 < \cdots < 2_{n_2} < \cdots < r_1 < \cdots < r_{n_r},$$

and the first column word is $1_1 \cdots 1_{n_1} 2_1 \cdots 2_{n_2} \cdots r_1 \cdots r_{n_r}$. We look at the last column word: if $w_{n,j}$ and $w_{n,j+1}$ are both some letter k, the order between these two occurrences of k is given by the next letter in the conjugates of w, and these are respectively $w_{1,j}$ and $w_{1,j+1}$. Thus $w_{n,j} < w_{n,j+1}$. Together with the hypothesis, this implies that the last column word is

$$(\pi 1)_1 \cdots (\pi 1)_{n_{\pi 1}} \cdots (\pi r)_1 \cdots (\pi r)_{n_{\pi r}}.$$

Thus, if we regard the rule $w_{1,j} \mapsto w_{n,j}$ as a mapping on the $n_1 + \ldots + n_r$ points

$$\{1_1, \ldots, 1_{n_1}, 2_1, \ldots 2_{n_2}, \ldots, r_1, \ldots, r_{n_r}\},$$

and put $\Delta_i = \{i_1, \ldots, i_{n_i}\}$, we get the claimed result. □

Corollary 7. *If the discrete r-interval exchange transformation T with length vector $(n_1, n_2, \ldots, n_r)$, and permutation π is not minimal, the word $(\pi 1)^{n_{\pi 1}} \ldots (\pi r)^{n_{\pi r}}$ has no primitive pre-image by the Burrows-Wheeler transform.*

Proof. Let w be such an antecedent. By the previous lemma, the map $w_{1,j} \mapsto w_{n,j}$ corresponds to T. If T is not minimal, there is a proper subset E of

$$\{1_1, \ldots, 1_{n_1}, 2_1, \ldots 2_{n_2}, \ldots, r_1, \ldots, r_{n_r}\}$$

which is invariant under $w_{1,j} \mapsto w_{n,j}$. Thus in the conjugates of w, preceding any occurrence of a letter of E is another occurrence of a letter of E. This implies that w is made up entirely of letters of E, a contradiction. □

Proof. ((1) implies (2)) Let w be as in the hypothesis. Then $B(w) = (\pi 1)^{n_{\pi 1}} \cdots (\pi r)^{n_{\pi r}}$. Thus the transformation T of Lemma 7 is minimal, and thus has a periodic trajectory $w'w'w' \ldots$, where w' has Parikh vector $(n_1, \ldots, n_r)$. If $w' = u^k$, then $n_i = kn'_i$ for all i, and the set made with the n'_i leftmost points of each Δ_i is T-invariant, thus w' must be primitive.

By the proof, made above, that (2) implies (1), w' is π-clustering. Hence $B(w') = B(w)$ and, by Lemma 5, w is conjugate to w', hence ww occurs also in a trajectory of T. □

Some of the hypotheses of Theorem 4 may be weakened.

Alphabet. $\{1, \ldots, r\}$ can be replaced by any ordered set $A = \{a_1 < a_2 < \cdots < a_r\}$ by using a letter-to-letter morphism. Thus for a given word w, we can restrict the alphabet to the letters occurring in w. Note that if ww occurs in a trajectory of an r-interval exchange transformation, but only the letters $j_1, \ldots, j_d$ occur in w, then, by the reasoning of the proof that (4) implies (1), w is π'-clustering, where π' is the unique permutation on $\{1, \ldots, d\}$ such that $(\pi')^{-1}(y) < (\pi')'^{-1}(z)$ iff $\pi^{-1}(j_y) < \pi^{-1}(j_z)$. If π is a permutation defining perfect clustering, then so is π'.

Primitivity. The Burrows-Wheeler transformation can be extended to a non-primitive word $w_1 \cdots w_n$, by ordering its n (non necessarily different) conjugates $w_i \cdots w_n w_1 \cdots w_{i-1}$ by non-strictly increasing lexicographical order and taking the word made by their last letters.

In this case the result of Lemma 7 does not extend: For example $B(1322313223) = 3333222211$ though the discrete 3-interval exchange transformation with length vector $(2, 2, 4)$, and permutation $\pi 1 = 3, \pi 2 = 2, \pi 3 = 1$ is not minimal. Note that if $(\pi 1)^{n_{\pi 1}} \cdots (\pi r)^{n_{\pi r}}$ has a non-primitive antecedent by the Burrows-Wheeler transform, then the n_i have a common factor k. There exist (see below) non-minimal discrete interval exchange transformations which do not satisfy that condition, and thus words such as 32221 which have no antecedent at all by the Burrows-Wheeler transformation.

But our Theorem 4 is still valid for non-primitive words: the proof in the first direction does not use the primitivity, while in the reverse direction we write $w = u^k$, apply our proof to the primitive u, and check that u^{2k} occurs also in a trajectory.

Two permutations. An extension of Theorem 4 which fails is to consider, as the dynamicians do [16], interval exchange transformations defined by permutations π and π'; this amounts to coding the interval Δ_i by $\pi' i$ instead of i. A simple counter-example will be clearer than a long definition: take points $x_1, \ldots, x_9$ labelled 223331111 and send them to 111133322 by a (minimal) discrete 3-interval exchange transformation, but where the points are not labelled as in Definition 3 (namely $Tx_1 = x_8$, $Tx_3 = x_5$ etc...). Then $w = 123131312$ is such that ww occurs in trajectories of T but $B(w) = 323311112$.

4 Building clustering words

Theorem 4 provides two different ways to build clustering words, from infinite trajectories either of discrete (or rational) interval exchange transformations or of continuous aperiodic interval exchange transformations. For $r = 2$ and the permutation $\pi 1 = 2, \pi 2 = 1$, the first way gives all the periodic balanced words, and the second way gives (by Proposition 10 below) all infinite Sturmian words: both ways of building clustering words on two letters are used, explicitly or implicitly, in Jenkinson and Zamboni [8].

The use of discrete interval exchange transformations leads naturally to the question of characterizing all minimal discrete r-interval exchange transformations through their length vector; this has been solved by Pak and Redlich [13] for $n = 3$ and $\pi 1 = 3, \pi 2 = 2, \pi 3 = 1$: if the length vector is (n_1, n_2, n_3), minimality is equivalent to $(n_1 + n_2)$ and $(n_2 + n_3)$ being coprime. Thus

Example 8. With the discrete interval exchange $11223333 \rightarrow 333322111$, we get the perfectly clustering word 313131223.

The same reasoning can be extended to other permutations: for $\pi 1 = 2, \pi 2 = 3, \pi 3 = 1$, minimality is equivalent to n_1 and $(n_2 + n_3)$ being coprime; for $\pi 1 = 3, \pi 2 = 1, \pi 3 = 2$, minimality is equivalent to n_3 and $(n_2 + n_1)$ being coprime; for other permutation on these three letters, T is never minimal.

For $r \geq 4$ intervals, the question is still open. An immediate equivalent condition for non-minimality is $\sum_{i=1}^{m} s_{w_i} = 0$ for $m < n_1 + \cdots + n_r$ and $w_1 \cdots w_m$ a word occurring in a trajectory. It is easy to build non-minimal examples satisfying such an equality for simple words w, for example for $r = 4$ and $\pi 1 = 4, \pi 2 = 3, \pi 3 = 2, \pi 4 = 1$, $n_1 = n_2 = n_3 = 1$ gives non-minimal examples for any value of n_4, the equality being satisfied for $w = 24^q$ if $n_4 = 3q$, $w = 14^{q+1}$ if $n_4 = 3q + 1$, $w = 34^q$ if $n_4 = 3q + 2$. Similarly, the following example shows how we still do get clustering words, but they may be somewhat trivial.

Example 9. The discrete interval exchange $111233444 \to 444332111$ satisfies the above equality for $w = 14$; it is non-minimal and gives two perfectly clustering words on smaller alphabets, 41 and 323.

To study continuous aperiodic interval exchange transformations we need a technical condition called *i.d.o.c.* [9] which states that *the orbits of the discontinuities of T are infinite and disjoint*. It is proved in Keane [9] or in Viana [16] that this condition implies aperiodicity and minimality, and that, if π is *primitive*, i.e., $\pi\{1, \ldots, d\} \neq \{1, \ldots, d\}$ for $d < r$, then the r-interval exchange transformation with probability vector $(\alpha_1, \ldots, \alpha_r)$ and permutation π satisfies the i.d.o.c. condition if $\alpha_1, \ldots, \alpha_r$ and 1 are rationally independent. We can now prove

Proposition 10. *Let $w = w_1 \cdots w_n$ be a primitive word on $A = \{1, \ldots, r\}$, such that every letter of A occurs in w. Then w is π-clustering if and only if ww occurs in a trajectory of a continuous r-interval exchange transformation with permutation π, satisfying the i.d.o.c. condition.*

Proof. The "if" direction is as in Theorem 4. To get the "only if", we generate w by a minimal discrete interval exchange transformation as in (2) of Theorem 4, and thus π is primitive. Then we replace it by a continuous periodic interval exchange transformation as in the proof that (3) implies (4). But, because cylinders are always semi-open intervals, if a given word ww occurs in a trajectory of a continuous r-interval exchange transformation with permutation π and probability vector $(\alpha_1, \ldots, \alpha_r)$, it occurs also in trajectories of every r-interval exchange transformation with the same permutation whose probability vector is close enough to $(\alpha_1, \ldots, \alpha_r)$. Thus we can change the α_i to get the irrationality condition which implies the i.d.o.c. condition. □

Trajectories of interval exchange transformations satisfying the i.d.o.c. condition may be explicitly constructed via the *self-dual induction* algorithms of [5] for $r = 3$ and $\pi 1 = 3, \pi 2 = 2, \pi 3 = 1$, [6] for all r and $\pi i = r + 1 - i$, and the forthcoming [4] in the most general case. More precisely, Proposition 4.1 of [6] shows that if the permutation is $\pi i = r + 1 - i$ (or more generally if the permutation is in the *Rauzy class* of $\pi i = r + 1 - i$), then there exist infinitely many words ww in the trajectories. It also gives a sufficient condition for building such words: if a *bispecial* word w, a suffix s and a prefix p of w are such that $pw = ws$, then both pp and ss occur in the trajectories. In turn, a recipe to achieve that relation is given in (i) of Theorem 2.8 of [6]: we just need that in the underlying algorithm described in Section 2.6 of [6], either $p_n(i) = i$ or $m_n(i) = i$ (except for some initial values of n, where, for $i = 1$, p and s are longer than w). Many explicit examples of ww have been built in this way.

- For $r = 3$, $w = A_k$, $w = B_k$ (see Ferenczi, Holton and Zamboni [5, Prop. 2.10]),

Example 11. 13131312222 and 131312221312213122 are perfectly clustering.

- For $r = 4$, $w = M_2(k)$, $w = P_3(k)M_1(k)$ (see Ferenczi and Zamboni [7, Lemmas 4.1 and 5.1]),

Example 12. $2^m(3141)^n32$ are perfectly clustering for any m and n.

- For all $r = n$, $w = P_{k,1,1}$, $w = P_{k,n-i,i+1}P_{k,i+1,n-i}$, $w = M_{k,n+1-i,i-1}M_{k,i-1,n+1-i}$ (see Ferenczi [3, Theorem 12]);

Example 13. 5252434252516152516161525161 is perfectly clustering.

For other permutations, we describe in Ferenczi [4] an algorithm generalizing the one in Ferenczi and Zamboni [6], but we do not know if every interval exchange transformation produces infinitely many ww. For the permutation $\pi 1 = 4, \pi 2 = 3, \pi 3 = 1, \pi 4 = 2$, examples can be found in Theorem 5.2 of [6], with $w = P_{1,q_n}M_{2,q_n}$, $w = P_{2,q_n}M_{3,q_n}$, $w = P_{3,q_n}M_{1,q_n}$,

Example 14. 4123231312412 is π-clustering,

We remark that our self-dual induction algorithms for aperiodic interval exchange transformations generate families of nested clustering words with increasing length, and thus may be more efficient in producing very long clustering words than the more immediate algorithm using discrete interval exchange transformations.

References

[1] M. Burrows and D. J. Wheeler, A block sorting data compression algorithm, Technical report, Digital System Research Center, 1994.

[2] M. Crochemore, J. Désarménien, and D. Perrin, A note on the Burrows-Wheeler transformation, *Theoret. Comput. Sci.* **332** (2005), 567–572.

[3] S. Ferenczi, Billiards in regular $2n$-gons and the self-dual induction, *J. London Math. Soc.*, 2013, doi:10.1112/jlms/jds075. Preprint available at http://iml.univ-mrs.fr/~ferenczi/ngon.pdf.

[4] S. Ferenczi, The self-dual induction for every interval exchange transformation, preprint, 2012, available at http://iml.univ-mrs.fr/~ferenczi/fie.pdf.

[5] S. Ferenczi, C. Holton, and L. Q. Zamboni, Structure of three-interval exchange transformations. II. A combinatorial description of the trajectories, *J. Anal. Math.* **89** (2003), 239–276.

[6] S. Ferenczi and L. Q. Zamboni, Structure of $k-$interval exchange transformations: induction, trajectories, and distance theorems, *J. Anal. Math.* **112** (2010), 289–328.

[7] S. Ferenczi and L. Q. Zamboni, Eigenvalues and simplicity of interval exchange transformations, *Ann. Sci. Éc. Norm. Supér. (4)* **44** (2011), 361–392.

[8] O. Jenkinson and L. Q. Zamboni, Characterisations of balanced words via orderings, *Theoret. Comput. Sci.* **310** (2004), 247–271.

[9] M. Keane, Interval exchange transformations, *Math. Z.* **141** (1975), 25–31.

[10] S. Mantaci, A. Restivo, G. Rosone, and M. Sciortino, An extension of the Burrows-Wheeler transform, *Theoret. Comput. Sci.* **387** (2007), 298–312.

[11] S. Mantaci, A. Restivo, and M. Sciortino, Burrows-Wheeler transform and Sturmian words. *Inform. Process. Lett.* **86** (2003), 241–246.

[12] V. Oseledets, On the spectrum of ergodic automorphisms, *Doklady Akademii Nauk. SSSR* **168** (1966), 1009–1011. In Russian. English translation in *Soviet Math. Doklady* **7** (1966), 776–779.

[13] I. Pak and A. Redlich, Long cycles in abc-permutations, *Funct. Anal. Other Math.* **2** (2008), 87–92.

[14] A. Restivo and G. Rosone, Burrows-Wheeler transform and palindromic richness, *Theoret. Comput. Sci.* **410** (2009), 3018–3026.

[15] A. Restivo and G. Rosone, Balancing and clustering of words in the Burrows-Wheeler transform, *Theoret. Comput. Sci.* **412** (2011), 3019–3032.

[16] M. Viana, Dynamics of interval exchange maps and Teichmüller flows, preliminary manuscript available from http://w3.impa.br/~viana/, 2012.

2010 *Mathematics Subject Classification*: Primary 68R15.
Keywords: Burrows-Wheeler transform, discrete interval exchange transformation.

Received April 6 2012; revised version received August 23 2012. Published in *Journal of Integer Sequences*, March 2 2013.

Return to Journal of Integer Sequences home page.