Maximal unbordered factors of random strings

Cording, Patrick Hagge; Gagie, Travis; Knudsen, Mathias Bæk Tejs; Kociumaka, Tomasz

Published in:
Theoretical Computer Science

DOI:
10.1016/j.tcs.2020.11.019

Publication date:
2021

Document version
Early version, also known as pre-print

Citation for published version (APA):
Cording, P. H., Gagie, T., Knudsen, M. B. T., & Kociumaka, T. (2021). Maximal unbordered factors of random strings. *Theoretical Computer Science, 852*, 78-83. https://doi.org/10.1016/j.tcs.2020.11.019
Maximal Unbordered Factors of Random Strings

Patrick Hagge Cording$^1$, Travis Gagie$^{12}$, Mathias Bæk Tejs Knudsen$^{13}$, and Tomasz Kociumaka$^4$

$^1$DTU Compute, Technical University of Denmark, patrick.cording@gmail.com
$^2$CeBiB, EIT, Universidad Diego Portales, Chile, travis.gagie@gmail.com
$^3$Department of Computer Science, University of Copenhagen, Denmark, mathias@tejs.dk
$^4$Institute of Informatics, University of Warsaw, Poland, kociumaka@ mimuw.edu.pl

Abstract

A border of a string is a non-empty prefix of the string that is also a suffix of the string, and a string is unbordered if it has no border other than itself. Loptev, Kucherov, and Starikovskaya [CPM 2015] conjectured the following: If we pick a string of length $n$ from a fixed non-unary alphabet uniformly at random, then the expected maximum length of its unbordered factors is $n - O(1)$. We confirm this conjecture by proving that the expected value is, in fact, $n - \Theta(n^{-1})$, where $\sigma$ is the size of the alphabet. This immediately implies that we can find such a maximal unbordered factor in linear time on average. However, we go further and show that the optimum average-case running time is in $\Omega(\sqrt{n})/O(\sqrt{n \log n})$ due to analogous bounds by Czumaj and Gąsieniec [CPM 2000] for the problem of computing the shortest period of a uniformly random string.

1 Introduction

Let $\Sigma$ be a finite alphabet of size $\sigma \geq 2$. A string $S \in \Sigma^n$ is a sequence $S = S[1] \cdots S[n]$ of $n$ symbols from $\Sigma$; the length $n$ of $S$ is denoted by $|S|$. For $1 \leq i \leq j \leq n$, we denote $S[i, j] = S[i] \cdots S[j]$ and call the string $S[i, j]$ a factor of $S$. A factor $S[1, j]$ is a prefix of $S$ and a factor $S[i, n]$ is a suffix of $S$. A border of a string is a non-empty prefix of the string that is also a suffix of the string. In other words, the string $S$ has a border of length $\ell$, $1 \leq \ell \leq n$, if and only if $S[1, \ell] = S[n - \ell + 1, n]$.

A string $S$ is unbordered if it does not have any proper border, i.e., any border other than the whole of $S$. By $L(S)$ we denote the maximum length of unbordered factors of $S$. Any unbordered factor of length $L(S)$ is called a maximal unbordered factor of $S$.

An integer $p > 0$ is a period of a string $S \in \Sigma^n$ if $S[i] = S[i + p]$ for $1 \leq i \leq n - p$. The shortest period of a string $S$ is denoted $\text{per}(S)$. Note that $p$ is a period of $S$ if and only if $S$ has a border of length $n - p$, so $S$ is unbordered if and only if $\text{per}(S) = n$. Moreover, $\text{per}(S[i, j]) \leq \text{per}(S)$; applied to a maximal unbordered factor, this yields $L(S) \leq \text{per}(S)$.

Example 1 ([1]). If $S = 1011001101$, then $\text{per}(S) = 7$ and $L(S) = 6$. The maximal unbordered factors are $S[1, 6] = 101100$ and $S[5, 10] = 001101$.

Unbordered factors were first studied by Ehrenfeucht and Silberger [6], with emphasis on the relationship $\text{per}(S)$ and $L(S)$. The question when $\text{per}(S) = L(S)$ received more attention in the literature [1, 5, 9, 8]. For strings $S \in \Sigma^n$, the equality holds if $L(S) \leq \frac{3}{4}n$ [9] or $\text{per}(S) \leq \frac{1}{2}n$ [6].

*A preliminary version of this paper [3] with weaker results was presented at the 23rd Symposium on String Processing and Information Retrieval (SPIRE ’16).

$^1$Supported by the Danish Research Council under the Sapere Aude Program (DFF 4005-00267).

$^2$Supported by Fondecyt grant 1171058.

$^3$Partly supported by Mikkel Thorup’s Advanced Grant from the Danish Council for Independent Research under the Sapere Aude research career programme and the FNU project AlgoDisc — Discrete Mathematics, Algorithms, and Data Structures.
Loptev, Kucherov, and Starikovskaya [15] proved that for uniformly random string \( S \in \Sigma^n \) over an alphabet \( \Sigma \) of size \( \sigma \geq 2 \) the expected maximum length \( E[L(S)] \) of unbordered factors is at least \( n(1 - \xi(\sigma) \cdot \sigma^{-4}) + O(1) \), where \( \xi(\sigma) \) converges to 2 as \( \sigma \) grows. When \( \sigma \geq 5 \) and \( n \) is sufficiently large, their bound implies \( E[L(S)] \geq 0.99n \). Supported by experimental results, Loptev et al. [15] conjectured that \( E[L(S)] = n - O(1) \).

In Section 2, we confirm this conjecture and prove that the tail of \( n - L(S) \) decays exponentially.

**Theorem 2.** Let \( S \in \Sigma^n \) be a uniformly random string over an alphabet \( \Sigma \) of size \( \sigma \geq 2 \).

(a) \( E[L(S)] = n - O(\sigma^{-1}) \).

(b) For each \( \delta > 0 \), the probability of \( L(S) = n - O(\log_{\sigma} \delta^{-1}) \) is at least \( 1 - \delta \).

One can easily deduce that \( \text{per}(S) \geq L(S) \) also satisfies both claims of Theorem 2. However, a recent study by Holub and Shallit [10] provides much stronger results concerning the shortest periods of uniformly random strings.

The problem of computing a maximal unbordered factor of a uniformly random string was studied by Loptev et al. [15] and Gawrychowski et al. [7], who gave algorithms with average-case running times of \( O(\frac{n^2}{\sigma} + n) \) and \( O(n \log n) \), respectively. The solution by Loptev et al. [15, Theorem 3] actually takes \( O(n(n - L(S) + 1)) \) worst-case time. By Theorem 2(a), its average-case running time is therefore \( O(n) \). Nevertheless, this is still much worse than what is necessary to compute the shortest period of a uniformly random string [4]. To address this issue, in Section 3 we develop a pair of reductions using Theorem 2(b) to show that computing \( L(S) \) and \( \text{per}(S) \) is equivalent with respect to the average-case running time.

**Theorem 3.** Let \( S \in \Sigma^n \) be a uniformly random string over an alphabet \( \Sigma \) of size \( \sigma \).

(a) The problem of computing \( L(S) \) can be reduced in \( O(\log_{\sigma} n) \) expected time to the problem of computing \( \text{per}(S') \) for a fixed factor \( S' \) of \( S \).

(b) The problem of computing \( \text{per}(S) \) can be reduced in \( O(1) \) expected time to the problem of computing \( L(S) \).

Consequently, the \( \Omega(\sqrt{n}) \) and \( O(\sqrt{n \log_{\sigma} n}) \) lower and upper bounds known for computing the shortest period of a uniformly random string, both due to Czumaj and Gąsieniec [4], carry over to computing a maximal unbordered factor of such a string.

**Corollary 4.** The problem of computing a maximal unbordered factor of a uniformly random string over an alphabet \( \Sigma \) of size \( \sigma \) takes \( O(\sqrt{n \log_{\sigma} n}) \) time on average, and this bound is within an \( O(\sqrt{\log_{\sigma} n}) \) factor of optimal.

Czumaj and Gąsieniec also conjectured that the optimum average-case running time of computing the shortest period is \( \Theta(\sqrt{n \log_{\sigma} n}) \); any resolution of this conjecture automatically transfers to maximal unbordered factors.

The worst-case running time we get from Theorem 3 and Czumaj and Gąsieniec’s work [4] is \( O(n^2) \). However, to obtain state-of-the-art running time both in the average case and in the worst case, we can dovetail our solution with any of the worst-case algorithms for computing a maximal unbordered factor. Gawrychowski et al. [7] gave such an algorithm with the running time \( O(n^{1.5}) \). Very recently, this has been improved [12] to \( O(n \log n \log^2 \log n) \) (and further to \( O(n \log n) \) if one allows Las Vegas randomization). Nevertheless, this is still slower than the \( O(n) \) time needed to compute the shortest period in the worst-case [16, 11].

Data structures for answering a period queries have also recently been developed. Such a query takes two indices \( i \) and \( j \) and the answer is the shortest period \( \text{per}(S[i, j]) \). Kociumaka et al. [14] developed a data structure of size \( O(n) \) answering period queries in \( O(\log n) \) time, which improved upon several earlier time-space trade-offs they presented in an earlier paper [13]. Computing \( L(S[i, j]) \) for a given factor \( S[i, j] \) appears to be a much more difficult task.

Another interesting possibility is to extend our results from average-case analysis to smoothed analysis [17, 18, 2], in which the input can be chosen adversarially but some random noise is then added to it. We
conjecture that when the noise level is reasonably large — e.g., each symbol is replaced by a randomly chosen one with some positive constant probability — then our bounds do not change significantly. Our results or techniques could also be applicable to other problems concerning borders and periods.

2 Distribution of Maximum Length of Unbordered Factors

Let us fix an alphabet $\Sigma$ of size $\sigma \geq 2$. For every $n \geq 0$, we define a random variable $\Delta_n$ distributed as $|S| - L(S)$ for uniformly random $S \in \Sigma^n$. The following lemma, which gives a common upper bound of the moment-generating functions $M_{\Delta_n}(t) = E[e^{t\Delta_n}]$, is the key tool behind Theorem 2.

**Lemma 5.** For $n \in \mathbb{N}$ and $0 \leq t \leq 0.1 \ln \sigma$, we have $M_{\Delta_n}(t) \leq C(t)$, where

$$C(t) = \frac{\sigma^3 - 2\sigma^2 e^{2t} + e^{4t}}{\sigma^3 - 2\sigma^2 e^{2t} + e^{4t}}.$$

**Proof.** We proceed by induction on $n$. The base case is $n \in \{0, 1\}$ for which $\Delta_n = 0$ and therefore $M_{\Delta_n}(t) = 1$. Consequently, we need to prove that

$$C(t) - M_{\Delta_n}(t) = \frac{\sigma^3 - 2\sigma^2 e^{2t}}{\sigma^3 - 2\sigma^2 e^{2t} + e^{4t}} - 1 = \frac{e^{4t}}{\sigma^3 - 2\sigma^2 e^{2t} + e^{4t}} \geq 0.$$

Note that the denominator is a quadratic function of $e^{2t}$ with a minimum at $e^{2t} = \sigma^2$. Hence, $\sigma^3 - 2\sigma^2 e^{2t} + e^{4t} \geq 3 - 2\sigma^2 + \sigma^4$ for $t \leq 0.1 \ln \sigma$. The right-hand side is a polynomial of $\sigma^0.2$, and one can easily verify that it is positive for $\sigma \geq 2$. Consequently, the denominator is positive. To complete the proof of the base case, observe that $e^{2t}(\sigma^2 - e^{2t})$ is also positive for $t \leq \ln \sigma$.

For $n \geq 2$, we assume $M_{\Delta_m}(t) \leq C(t)$ for $m < n$ and $0 \leq t \leq 0.1 \ln \sigma$. We consider a uniformly random $S \in \Sigma^n$ and condition over the possible lengths $\ell$ of the shortest border of $S$. More formally, we define $F(S)$ as the smallest integer $\ell > 0$ such that $S[1, \ell] = S[n - \ell + 1, n]$, and we write

$$M_{\Delta_n}(t) = E[e^{t(n - L(S))}] = \sum_{\ell=1}^{n} P[F(S) = \ell] \cdot E[e^{t(n - L(S))} | F(S) = \ell].$$

Now, we bound from above individual terms of this sum. Observe that $F(S) = n$ is equivalent to $L(S) = n$ and therefore

$$E[e^{t(n - L(S))} | F(S) = n] = 1. \tag{3}$$

For $\ell \leq \frac{1}{2}n$, we observe that $S[\ell + 1, n - \ell]$ is independent from $F(S) = \ell$. Due to $L(S) \geq L(S[\ell + 1, n - \ell])$, this yields

$$E[e^{t(n - L(S))} | F(S) = \ell] \leq E[e^{t(n - L(S[\ell + 1, n - \ell]))} | F(S) = \ell] = E[e^{t(n - 2\ell - L(S[\ell + 1, n - \ell]))}] = e^{2\ell t} M_{\Delta_{n-2\ell}}(t). \tag{4}$$

Moreover, we note that $F(S) = \ell$ implies $S[i] = S[n - \ell + i]$ for $1 \leq i \leq \ell$ and these events are independent. For $\ell \geq 2$, we have one more independent event $S[1] \neq S[\ell]$ due to $F(S) \neq 1$. Consequently,

$$P[F(S) = \ell] \leq \begin{cases} \sigma^{-1} & \text{if } \ell = 1, \\ (\sigma - 1)^{\ell-1} & \text{if } 2 \leq \ell \leq \frac{1}{2}n. \end{cases} \tag{5}$$

In the remaining case of $\frac{1}{2}n < \ell < n$, we observe that if $S[1, \ell] = S[n - \ell + 1, n]$, then $S[n - \ell + 1, \ell]$ is also a border of $S$. This contradicts $F(S) = \ell$ because $|S[n - \ell + 1, \ell]| = 2\ell - n < \ell$. Consequently,

$$P[F(S) = \ell] = 0 \quad \text{if } \frac{1}{2}n < \ell < n. \tag{6}$$
Plugging (3–6) into (2), we obtain

\[ M_{\Delta_n}(t) \leq P[F(S) = n] + \sum_{\ell=1}^{\lfloor n/2 \rfloor} P[F(S) = \ell] \cdot e^{2\ell t} \cdot M_{\Delta_n-\ell}(t) \]

\[ \leq 1 + \sigma^{-1} \cdot e^{2t} \cdot M_{\Delta_{n-2}}(t) + \sum_{\ell=2}^{\lfloor n/2 \rfloor} (\sigma - 1) \sigma^{-\ell-1} \cdot e^{2\ell t} \cdot M_{\Delta_{n-2\ell}}(t). \]  

(7)

The inductive assumption further yields

\[ M_{\Delta_n}(t) \leq 1 + \sigma^{-1} \cdot e^{2t} \cdot C(t) + \sum_{\ell=2}^{\lfloor n/2 \rfloor} (\sigma - 1) \sigma^{-\ell-1} \cdot e^{2\ell t} \cdot C(t) \]

\[ \leq 1 + C(t) \left( \sigma^{-1} e^{2t} + (\sigma - 1) \sigma^{-3} e^{4t} \right) \sum_{\ell=0}^{\infty} (\sigma^{-1} e^{2t})^\ell \]

\[ = 1 + C(t) \left( \sigma^{-1} e^{2t} + (\sigma - 1) \sigma^{-3} e^{4t} \cdot \frac{1}{1 - \sigma^{-1} e^{2t}} \right) \]

\[ = 1 + C(t) \cdot \frac{\sigma(\sigma - e^{2t})e^{2t} - (\sigma - 1)e^{4t}}{\sigma^2(\sigma - e^{2t})} \]

\[ = 1 + \frac{\sigma^3 - \sigma^2 e^{2t} - \sigma^2 e^{2t} - e^{4t}}{\sigma^3 - \sigma^4 e^{2t} + e^{4t}} \]

\[ = \frac{\sigma^3 - \sigma^2 e^{2t} + e^{4t} \sigma - e^{4t}}{\sigma^3 - \sigma^4 e^{2t} + e^{4t}} \]

\[ = C(t). \]  

(8)

This completes the proof of Lemma 5.

Next, let us focus on the expected value \( E[\Delta_n] \). Note that \( M_{\Delta_n}(t) = E[e^{t\Delta_n}] \geq E[1 + t\Delta_n] \). Consequently, for \( 0 < t \leq 0.1 \ln \sigma \) we have

\[ E[\Delta_n] \leq \frac{M_{\Delta_n}(t) - 1}{t} \leq \frac{C(t) - 1}{t}. \]  

(9)

Hence, \( E[\Delta_n] \) is bounded by a function of \( \sigma \) independent of \( n \). To analyze its asymptotics in terms of \( \sigma \), we plug \( t = 1 \) (valid for \( \sigma \geq e^{10} \)), which yields

\[ E[\Delta_n] \leq C(1) - 1 = \frac{\sigma^2 e^2 - e^4}{\sigma^3 - 2\sigma^2 e^2 + e^4} = \frac{O(\sigma^2)}{\Omega(\sigma^3)} = O(\sigma^{-1}). \]  

(10)

This completes the proof of Theorem 2(a).

For the claim (b), we apply Markov’s inequality on top of Lemma 5:

\[ P[\Delta_n \geq \ell] \leq \frac{E[e^{\ell \Delta_n}]}{e^{\ell t}} = \frac{M_{\Delta_n}(t)}{e^{\ell t}} \leq \frac{C(t)}{e^{\ell t}}. \]  

(11)

Hence, it suffices to take \( \ell \geq 10 \log_{\sigma}(\delta^{-1} \cdot C(0.1 \ln \sigma)) \) to make sure that the probability does not exceed \( \delta \). To complete the proof, observe that

\[ C(0.1 \ln \sigma) = \frac{\sigma^3 - \sigma^{2.2}}{\sigma^3 - 2\sigma^2 + \sigma^{0.4}} = \frac{O(\sigma^3)}{\Omega(\sigma^3)} = O(1). \]  

(12)
3 Average-Case Algorithms for Maximal Unbordered Factors

In this section, we give a pair of reductions between the problems of computing the shortest period and the maximum length of unbordered factors of a uniformly random string, thereby proving Theorem 3. We assume that the alphabet $\Sigma$ is of size $\sigma \geq 2$. Otherwise, both values are always 1.

We start with a simple argument showing Theorem 3(b). Suppose that we aim at computing $\operatorname{per}(S)$ for a uniformly random string $S \in \Sigma^n$. Having determined $L(S)$, we rely on the fact that $\operatorname{per}(S) \leq L(S)$. We construct a string $S_\delta := S[1, n - L(S)] \$ \$ S[L(S) + 1, n]$, where $\$ \$ \notin \Sigma$ is a sentinel symbol, and observe that $S$ has a border of length $\ell \leq n - L(S)$ if and only if $S_\delta$ has such a border. Moreover, the presence of the sentinel symbol guarantees that $S_\delta$ does not have proper borders longer than $n - L(S)$. Consequently, we have $|S| - \operatorname{per}(S) = |S_\delta| - \operatorname{per}(S_\delta)$. The value $\operatorname{per}(S_\delta)$ can be computed using a worst-case algorithm [16, 11], which takes $O(|S_\delta|) = O(n - L(S) + 1)$ time. The expected running time of the reduction is $O(1)$ due to Theorem 2(a).

We proceed with a proof of Theorem 3(a). Suppose that we aim at computing $L(S)$ for a uniformly random string $S \in \Sigma^n$. We apply Theorem 2(b) for $\delta = \frac{1}{n^2}$ to obtain a value $d = O(\log_\sigma n)$ such that $P[|T| - L(T) \geq d] \leq \frac{1}{n^2}$ for uniformly random strings $T \in \Sigma^m$ of arbitrary length $m$. Note that this also yields $P[|T| - \operatorname{per}(T) \geq d] \leq \frac{1}{n^2}$ due to $\operatorname{per}(T) \geq L(T)$.

If $n \leq 6d$, we simply determine $L(S)$ using Loptev et al.’s algorithm [15], which takes $O(d) = O(\log_\sigma n)$ time on average. Otherwise, we construct three strings

\[
\begin{align*}
\bar{S} & := S[1, 3d] S[n - 3d + 1, n], \\
S' & := S[d + 1, n - d], \\
\bar{S}' & := S[d + 1, 3d] S[n - 3d + 1, n - d],
\end{align*}
\]

and we compute $|\bar{S}| - L(\bar{S})$, $|S'| - \operatorname{per}(S')$, and $|\bar{S}'| - \operatorname{per}(\bar{S}')$. If any of these values exceeds $d$, we fall back to the algorithm of [15] to compute $L(S)$. Otherwise, we determine $L(S)$ based on $|\bar{S}| - L(\bar{S}) = |S| - L(|S|)$.

Before proving this equality, let us analyze the running time of the reduction. Observe that $S, S'$, and $\bar{S}'$ are uniformly random strings of the respective lengths, which lets us use average-case algorithms. In particular, it takes $O(d)$ time on average to compute $L(S')$ using Loptev et al.’s algorithm [15]. Determining $\operatorname{per}(S')$ is the target of the reduction, so we do not include it in the analysis. The value $\operatorname{per}(\bar{S}')$ is computed in $O(d)$ worst-case time [16, 11]. The probability of a fall-back is at most $\frac{1}{n^2}$ by the choice of $d$, which compensates for the worst-case\(^1\) time $O(n^2)$ it takes to apply Loptev et al.’s algorithm to the whole of $S$.

Overall, the reduction works in $O(d) = O(\log_\sigma n)$ time on average.

It remains to prove $|S| - L(S) = |\bar{S}| - L(\bar{S})$ provided that $|\bar{S}| - L(\bar{S}) \leq d$, $|S'| - \operatorname{per}(S') \leq d$, and $|\bar{S}'| - \operatorname{per}(\bar{S}') \leq d$. First, consider a maximal unbordered factor of $\bar{S}$. It must be of the form $S[i, 3d] S[n - 3d + 1, j]$ for some $1 \leq i \leq d$ and $n - d + 1 \leq j \leq n$, and we claim that $S[i, j]$ is then an unbordered factor of $S$. For a proof by contradiction, suppose that $S[i, j]$ has a proper border and the longest such border is of length $\ell$. Note that $\ell > \min(|S[i, 3d]|, |S[n - 3d + 1, j]|)$ because $S[i, 3d] S[n - 3d + 1, j]$ is unbordered. We conclude that $\operatorname{per}(S[i, j]) = |S[i, j]| - \ell < n - 3d$. However, this yields $\operatorname{per}(S') \leq \operatorname{per}(S[i, j]) < n - 3d = |S'| - d$, a contradiction. Consequently, $|S| - L(S) \leq |\bar{S}| - L(\bar{S})$.

The proof of $|S| - L(S) \geq |\bar{S}| - L(\bar{S})$ is symmetric. We consider a maximal unbordered factor $S[i, j]$ of $S$, observe that $1 \leq i \leq d$ and $n - d + 1 \leq j \leq n$ due to $|S| - L(S) \leq d$, and claim that $S[i, 3d] S[n - 3d + 1, j]$ is unbordered. For a proof by contradiction we suppose that it has a border of length $\ell$. We note that $\ell > \min(|S[i, 3d]|, |S[n - 3d + 1, j]|)$ because $S[i, j]$ is unbordered and derive $\operatorname{per}(S') \leq \operatorname{per}(S[i, 3d] S[n - 3d + 1, j]) < 3d$, which contradicts $\operatorname{per}(S') \geq |S'| - d = 3d$.

This completes the proof of Theorem 3(a).

\(^1\)Note that we cannot use the average-case bound of $O(n)$ because the conditional distribution of $S$ (in case of a fall-back) is no longer uniform across $\Sigma^n$.
Acknowledgments

Many thanks to Danny Hucke for asking about the possibility of a sublinear average-case algorithm at the presentation of the conference version of this paper, and to the anonymous reviewers for their comments.

References

[1] Roland Assous and Maurice Pouzet. Une caractérisation des mots periodiques. Discrete Mathematics, 25(1):1–5, 1979. doi:10.1016/0012-365X(79)90146-8.
[2] Christina Boucher and Kathleen Wilkie. Why large closest string instances are easy to solve in practice. In Edgar Chávez and Stefano Lonardi, editors, String Processing and Information Retrieval, SPIRE 2010, volume 6393 of LNCS, pages 106–117. Springer, 2010. doi:10.1007/978-3-642-16321-0_10.
[3] Patrick Hagge Cording and Mathias Bæk Tejs Knudsen. Maximal unbordered factors of random strings. In Shunsuke Inenaga, Kunihiko Sadakane, and Tetsuya Sakai, editors, String Processing and Information Retrieval, SPIRE 2016, volume 9954 of LNCS, pages 93–96, 2016. doi:10.1007/978-3-319-46049-9_9.
[4] Artur Czumaj and Leszek Gąsieniec. On the complexity of determining the period of a string. In Raffaele Giancarlo and David Sankoff, editors, Combinatorial Pattern Matching, CPM 2000, volume 1848 of LNCS, pages 412–422. Springer, 2000. doi:10.1007/3-540-45123-4_34.
[5] Jean-Pierre Duval. Relationship between the period of a finite word and the length of its unbordered segments. Discrete Mathematics, 40(1):31–44, 1982. doi:10.1016/0012-365X(82)90186-8.
[6] Andrzej Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. Discrete Mathematics, 26(2):101–109, 1979. doi:10.1016/0012-365X(79)90116-X.
[7] Paweł Gawrychowski, Gregory Kucherov, Benjamin Sach, and Tatiana Starikovskaya. Computing the longest unbordered substring. In Costas S. Iliopoulos, Simon J. Puglisi, and Emine Yilmaz, editors, String Processing and Information Retrieval, SPIRE 2015, volume 9309 of LNCS, pages 246–257. Springer, 2015. doi:10.1007/978-3-319-23826-5_24.
[8] Tero Harju and Dirk Nowotka. Periodicity and unbordered words: A proof of the extended Duval conjecture. Journal of the ACM, 54(4):20, 2007. doi:10.1145/1255443.1255448.
[9] Stepan Holub and Dirk Nowotka. The Ehrenfeucht–Silberger problem. Journal of Combinatorial Theory, Series A, 119(3):668–682, 2012. doi:10.1016/j.jcta.2011.11.004.
[10] Stepan Holub and Jeffrey Shallit. Periods and borders of random words. In Nicolas Ollinger and Heribert Vollmer, editors, Symposium on Theoretical Aspects of Computer Science, STACS 2016, volume 47 of LIPIcs, pages 44:1–44:10. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.STACS.2016.44.
[11] Donald E. Knuth, James H. Morris, Jr., and Vaughan R. Pratt. Fast pattern matching in strings. SIAM Journal on Computing, 6(2):323–350, 1977. doi:10.1137/0206024.
[12] Tomasz Kociumaka, Ritu Kundu, Manal Mohamed, and Solon P. Pissis. Longest unbordered factor in quasilinear time. In Seok-Hee Hong, editor, Algorithms and Computation, ISAAC 2018, volume 123 of LIPIcs, pages 70:1–70:13. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018. arXiv:1805.09924, doi:10.4230/LIPIcs.ISAAC.2018.70.
[13] Tomasz Kociumaka, Jakub Radoszewski, Wojciech Rytter, and Tomasz Waleń. Efficient data structures for the factor periodicity problem. In Liliana Calderón-Benavides, Cristina N. González-Caro, Edgar Chávez, and Nivio Ziviani, editors, \textit{String Processing and Information Retrieval, SPIRE 2012}, volume 7608 of \textit{LNCS}, pages 284–294. Springer, 2012. doi:10.1007/978-3-642-34109-0_30.

[14] Tomasz Kociumaka, Jakub Radoszewski, Wojciech Rytter, and Tomasz Waleń. Internal pattern matching queries in a text and applications. In Piotr Indyk, editor, \textit{26th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015}, pages 532–551. SIAM, 2015. doi:10.1137/1.9781611973730.36.

[15] Alexander Loptev, Gregory Kucherov, and Tatiana Starikovskaya. On maximal unbordered factors. In Ferdinando Cicalese, Ely Porat, and Ugo Vaccaro, editors, \textit{Combinatorial Pattern Matching, CPM 2015}, volume 9133 of \textit{LNCS}, pages 343–354. Springer, 2015. doi:10.1007/978-3-319-19929-0_29.

[16] James H. Morris, Jr. and Vaughan R. Pratt. A linear pattern-matching algorithm. Technical Report 40, Department of Computer Science, University of California, Berkeley, 1970.

[17] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. \textit{Journal of the ACM}, 51(3):385–463, 2004. doi:10.1145/990308.990310.

[18] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis: an attempt to explain the behavior of algorithms in practice. \textit{Communications of the ACM}, 52(10):76–84, 2009. doi:10.1145/1562764.1562785.