Unavoidable sets of constant length

A set X of words is unavoidable on A: any long enough word on A has a factor in X .

$$X = \{a, bbb\}$$

is unavoidable on $\{a, b\}$

$$X = \{aaa, aab, bab, bbb\}$$

also.

A finite set X is unavoidable iff every *periodic* infinite word has a factor in X.

Conjugacy classes

An unavoidable set of words of length n contains at least one word of every class of conjugacy of words of length n.

Let $c(k, n)$ be the number of conjugacy classes of words of length n on k symbols.

n	1	2	3	4	5	6	7	8	9	10
$c(n,2)$	2	3	4	6	8	14	20	36	60	108
$c(n,3)$	3	6	11	24	51	130	315	834	2195	5934
$c(n,4)$	4	10	24	70						
$c(n,5)$	5	10								

One has

$$c(n, k) = \frac{1}{n} \sum_{d|n} \varphi(n/d) k^d$$

where φ is the Euler function.

Minimal sets

Theorem (G. Hansel, J. M. Champarnaud, D.P., 2002)*For all* $n, k \geq 1$ *there exists an unavoidable set of* $c(k, n)$ *words of length* n *on* k *symbols.*

$n = 3, 4, 5, 6$

aaa	$aaaa$	$aaaaa$	$aaaaaa$
aab	$aaab$	$aaaab$	$aaaaab$
bab	$baab$	$baaab$	$baaaab$
bbb	$abab$	$abaab$	$abaaab$
	$babb$	$bbaab$	$bbaaab$
	$bbbb$	$babab$	$aabaab$
		$bbabb$	$baabab$
		$bbbbb$	$abaabb$
			$bbaabb$
			$ababab$
			$bbabab$
			$abbabb$
			$bbabbb$
			$bbbbbb$

Lyndon words

A Lyndon word is a primitive word minimal in its conjugacy class. Let P be the set of prefixes of Lyndon words. It is equal to the set of sesquipowers of Lyndon words= prefixes of powers of Lyndon words.

A *division* of a word $w \in P$ is a pair (l^p, u) such that $w = l^p u$ where $l \in L$, $p \geq 1$ and $u \in A^*$ with $|u| < |l|$.

Each word in P admits at least one division.

The *main division* of $w \in P$ is the division (l^p, u) where l is the shortest possible Lyndon word.

Let M_n be the set of minimal words of length n on k symbols.

The set

$$I_n = \{ul^n | (l^p, u) \text{ main division of } m \in M_n\}$$

is an unavoidable set.

$$n = 7$$

$aaaaaaa$	$aaaaaaa$
$aaaaaab$	$aaaaaab$
$aaaaabb$	$baaaaab$
$aaaabab$	$abaaaab$
$aaaabbb$	$bbaaaab$
$aaabaab$	$aabaaab$
$aaababb$	$abbaaab$
$aaabbab$	$babaaab$
$aaabbbb$	$bbbaaab$
$aabaabb$	$baabaab$
$aababab$	$abaabab$
$aababbb$	$bbaabab$
$aabbabb$	$abbaabb$
$aabbbab$	$babaabb$
$aabbbbb$	$bbbaabb$
$abababb$	$bababab$
$ababbbb$	$bbababb$
$abbabbb$	$babbabb$
$abbbbbb$	$bbbabbb$
$bbbbbbb$	$bbbbbbb$

Proof of the theorem

Proposition 1 *Let λ and l be two Lyndon words, with λ a prefix of l. Let $s \in A^*$ be a proper suffix of l, with $|s| < |\lambda|$. Then for all $n > 0$, the word $w = \lambda^n s$ is a Lyndon word.*

Proposition 2 *Let w be a prefix of a minimal word and let (λ^p, u) be its main division. Let $u' \in A^*$ be a word of the same length than u and such that the word $w' = \lambda^p u'$ is also a prefix of a minimal word. Then the main division of w' is the pair (λ^p, u').*

Proposition 3 *Let l be a Lyndon word. Let $p \in N$ be the smallest integer such that $|l^p| > n$. Then the word l^{p+1} has a factor in I_n.*