Tutorial Week 1

Definition 1. *For a* **string** w*, we say that* w *is* **primitive** *if there exists no other string* u *such that* $w = u^k$*, for some integer* $k > 1$*. Here* u *is called a* **root** *of* w*. Moreover, if* $w = uvz$ *for some strings* u, v, z*, then we call* u *a* **prefix** *of* w*,* z *a* **suffix** *of* w*, and each of* u, v, z *is called a* **factor**.

Exercise 1. *For the following list of strings, indicate their prefixes, suffixes, roots, and say if they are primitive or not.*

	prefixes	*suffixes*	*roots*	*primitive*
$ababab$				
$aaaaaa$				
$abcacb$				

Solution:

	prefixes	suffixes	roots	primitive
$ababab$	$\{a, ab, aba, abab, ababa, ababab\}$	$\{b, ab, bab, abab, babab, ababab\}$	$\{ab, ababab\}$	NO
$aaaaaa$	$\{a, aa, aaa, aaaa, aaaaa, aaaaaa$	$\{a, aa, aaa, aaaa, aaaaa, aaaaaa\}$	$\{a, aa, aaa, aaaaaa\}$	NO
$abcacb$	$\{a, ab, abc, abca, abcac, abcacb\}$	$\{b, cb, acb, cacb, bcacb, abcacb\}$	$\{abcacb\}$	YES

■

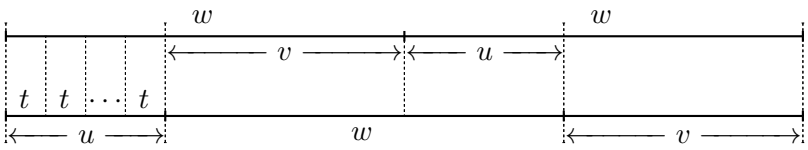

Figure 1: w has another occurrence in ww, neither as a prefix, nor as a suffix

Lemma 2. *If* w *is a primitive string, then* w *occurs as a factor of* ww *only as a prefix or a suffix.*

Proof. Assume this is not the case, and consider the shortest primitive string w that has a further occurrence as a factor of ww, neither as a prefix, nor as a suffix. Therefore, there exist non-empty strings u, v such that $ww = uwv$. It immediately follows that u is a prefix of w, while v is a suffix of w. Furthermore, since $2|w| = |u| + |w| + |v|$, we have that $w = uv$. The situation is depicted in Figure 1.

If we look at the alignment of $ww = uvuv$ with uwv, we immediately observe that $w = vu$, as well. Therefore, if $|u| = |v|$, we immediately have that $u = v$, thus $w = u^2 = v^2$, which contradicts the fact that w is primitive. Thus one of them must be longer than the other.

Assume without loss of generality that $|u| < |v|$ (the other case is symmetrical). Moreover, since every string can be expressed as a power of a primitive root, it must be that $u = t^k$, for some primitive string t and an integer $k > 0$. However, by looking at the alignment between $ww = uvuv = t^k v t^k v$ with $uwv = uuvv = t^k t^k vv$, we have that $vt^k = t^k v$. It is easy to observe in the figure that, in this case, it must be that for some prefix t' of t, and some integer $s \geq k$, we have $v = t^s t'$.

If t' is empty or $t' = t$, then it follows that both u and v are powers of t, and we conclude. Otherwise, looking at the suffix of length tt of vt^k and the corresponding suffix in $t^k v$, we note that the last occurrence of t in $v = t^s t'$ aligns with a factor of tt which is neither a prefix nor a suffix. Since $|t| \leq |u| < |w|$, and we assumed that w is the shortest primitive string for which this case occurs, the conclusion follows. □

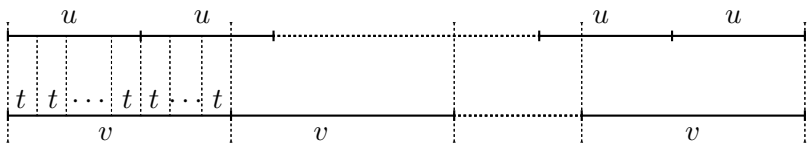

Figure 2: u^k and v^ℓ are in a perfect alignment, and $u = t^k$

Proposition 3. *If for two strings u, v we have that $u^k = v^\ell$, for some integers $k, \ell > 0$, then u and v are powers of the same string.*

Proof. Note that if u and v have the same length, then they are equal and the conclusion follows. Therefore, we can assume that one of them is longer than the other. Assume without loss of generality that $|u| < |v|$.

Since $u^k = v^\ell$, it follows that u is both a prefix of v, as well as a suffix of it. If $\ell = 1$, then the conclusion follows once more, as both strings are powers of u. Therefore, let us consider that $\ell > 1$.

If we denote by t the primitive root of u, then, for some positive integer r, we have $u = t^r$. Let s be the smallest integer such that $r < s < rk$ and $|t^s| > |v|$. Obviously, since $u^k = v^\ell$, such an integer exists. Figure 2 depicts this situation. However, in this case, (as can be seen from the Figure) we note that there exists a prefix t of the first occurrence of v that is a factor of tt, represented by the last two occurrences of t in t^s. Since t is primitive, it follows from Lemma 2 that t can only occur as a prefix of this element, and, therefore, as it can be easily seen from the Figure, $v = t^s$. □

Proposition 4. *There exists one and only one primitive root for every string.*

Proof. Assume that for some string w there exist two primitive roots u and v, such that $w = u^\ell$ and $w = v^k$, for some integers $\ell, k > 0$. However, in this case we have that $u^\ell = v^k$, and by Proposition 3, we have that u and v are powers of the same string. Hence, they are either equal, or one of them is not primitive. The later, however, is a contradiction of our assumption that both u and v are primitive, thus the conclusion regarding the uniqueness of the root follows. □

Exercise 5. *Considering the brute-force algorithm presented in the course, how many comparisons does the algorithm do in order to find all occurrences of the pattern* `ana` *in the text* `bananas`*? Please recall that the algorithm will stop only when it reaches the end of the text. What about when we look in the text* `anaeatsabanana`*? What about in the text* `anabanananananann`*?*

Solution: For the text `bananas` the brute-force algorithm does 9 comparisons in order to find all 2 occurrences of the pattern `ana`.

For the text `anaeatsabanana` the brute-force algorithm does 21 comparisons in order to find all 3 occurrences of the pattern `ana`.

For the text `anabanananananann` the brute-force algorithm does 26 comparisons in order to find all 5 occurrences of the pattern `ana`. ■