

1154-05-263

Andrzej Ehrenfeucht* (andrzej@cs.colorado.edu), 2736 Cheyenne Drive, Las Cruces, NM 88011. *Syntactic Tokens*.

Definition: The set of strings S over an alphabet A is a set of syntactic tokens if and only if (1) S is finite. (2) No two strings in S are substrings of each other. (3) S is closed under all permutations of letters in A . (4) Every string in A^* is covered by instances of tokens from S , with the possible exception of a prefix and suffix, each of which has a bounded length. (5) Let $t(w)$, where w is in A^* , be the sequence of uncovered prefix and suffix and all tokens from S in between. For every u, v in A^* , if $t(u) = t(v)$, then $u = v$. Definitions: A string is simple if and only if all its characters are different, and it is otherwise structured. A string awb , where a, b are in A , and w is in A^+ , is bounded if and only if w is simple and both aw and wb are structured. Theorem: The set of all bounded strings over alphabet A is a set of syntactic tokens. In the talk we introduce the concept of syntactic tokens, and describe combinatorial properties of sets of bounded strings over an alphabet of 20 letters. And we will suggest that sequences of tokens $t(w)$ may be used as tools for comparing and classifying primary structures of proteins, when A is the set of amino-acids. (Received August 27, 2019)