Division of Computer Science

Language Processing Systems Laboratory

Satoshi Okawa
Professor

Taro Suzuki
Senior Associate Professor
Summary of Achievement

Grants

Academic Activities
   member
   member
   member

Ph.D and Others Theses
   Thesis Advisor: T. Suzuki
   Thesis Advisor: T. Suzuki
   Thesis Advisor: T. Suzuki
   Thesis Advisor: T. Suzuki

Others
Summary of Achievement


This paper offers a new efficient regular expression matching algorithm which follows the leftmost-longest rule specified in POSIX 1003.2 standard. The algorithm basically emulates the subset construction without backtracking, so that its computational cost even in the worst case does not explode exponentially; the time complexity of the algorithm is $O(m(n^2 + c))$, where $m$ is the length of a given input string, $n$ the number of occurrences of the most frequently used letters in a given regular expression and $c$ the number of subexpressions to be used for capturing substrings plus the number of repetition operators. The correctness of the algorithm is given with respect to a formalization of the leftmost-longest semantics by means of a priority order on parse trees.