

実行時コンパイルを用いた正規表現評価器の実装

新屋 良磨

当研究室では, Concinnuation based C という状態遷移記述に適した C の下位言語を提案している. Continuous bsd C はステートメントより大きく, 関数よりも小さなプログラミング単位としてコードセグメントを持ち, コードセグメントからの継続を基本としている. 本研究では, 与えられた正規表現から, 等価な右舷状態オートマトンに変換し, さらにその状態遷移を CbC 等のプログラミング言語の記述に変換し, 実行時コンパイルによって得られた正規表現評価器を生成するコンパイラを Python で記述し, 評価を行った.

1 はじめに

...

2 CbC

...

3 正規表現

4 正規表現によるテキストマッチ

正規表現は与えられた文字列を受理するかどうかを判定できるパターンマッチングの機構であり, sed, grep, awk を始めとするテキスト処理ツールに広く利用されている. 正規表現には定められた文法によって記述され, 例えば, 正規表現 “a*b” は “a” の 0 回以上の繰り返し “b” で終わる文字列 (“b”, “ab”, “aaaab”) を受理し, “a(b—c)” は “a” で始まり, 直後が “b” または “c” で終わる文字列 (“ab”, “ac”) を受理する.

5 正規表現の評価器

6 grep

7 正規表現から CbC への変換

参考文献

- [1] Lanin, V. and Shasha, D.: A Symmetric Concurrent B-Tree Algorithm, Proc. 1986 Fall Joint Computer Conference, IEEE, 1986, pp. 380–389.
- [2] Sleator, D. D. and Tarjan, R. E.: Self-Adjusting Binary Search Trees, *J. ACM*, Vol. 32, No. 3 (1985), pp. 652–686.
- [3] Shapiro E.: The Family of Concurrent Logic Programming Languages. *ACM Computing Surveys*, Vol. 21, No. 3 (1989), pp. 413–510.
- [4] Tarjan, R. E.: Amortized Computational Complexity, *SIAM J. Alg. Disc. Math.*, Vol. 6, No. 2 (1985), pp. 306–318.
- [5] 和田久美子: スプレイ木の並列データ探索, Proc. KLI Programming Workshop '90, Tokyo, ICOT, 1990, pp. 42–49.

A 付録: L^AT_EX による論文作成のガイド

ここに, 以前の `sample.tex` では, 論文作成のガイドがあったが, その内容は `guide.tex` に移動した.

Implimentation Regular Expression Engine with Just-In-Time Compile.

Shinya Ryoma, 琉球大学工学部情報工学学科, Dept. of Information Engineering, Ryukyu University.

コンピュータソフトウェア, Vol.16, No.5 (1999), pp.78–83.

[研究論文] 1999 年 8 月 3 日受付.