String Suffix Automata and Subtree Pushdown Automata

JAN JANOUŠEK

JAN.JANOUSEK@FIT.CVUT.CZ

DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF INFORMATION TECHNOLOGIES
CZECH TECHNICAL UNIVERSITY
MOTIVATION

STRINGOLOGY
BASIC TOOL: FINITE AUTOMATA
MOTIVATION

LINEARISED NOTATIONS OF TREES: PREFIX OR POSTFIX NOTATION

**Theorem 1.** Given a tree $t$ and its prefix notation $\text{pref}(t)$, all subtrees of $t$ in prefix notation are substrings of $\text{pref}(t)$. 
REGULAR TREE LANGUAGES are accepted by **FINITE TREE AUTOMATA**.

**DETERMINISTIC PUSHDOWN AUTOMATA** accept a proper superclass of the regular tree languages in prefix or postfix notation.
MOTIVATION

STRINGOLOGY

BASIC TOOL: FINITE AUTOMATA

ARBOLGY – TREE ALGORITHMS, FOUNDED 2008
BASIC TOOL: (DETERMINISTIC) PUSHDOWN AUTOMATA

ARBOLGY AS AN ANALOGY WITH STRINGOLOGY

WEB PAGES http://www.arbology.org
PREFIX NOTATION

GRAMMAR FOR TREE WITH NODES HAVING RANK 0, 1, 2:

1. $S \rightarrow a_0$
2. $S \rightarrow a_1 S$
3. $S \rightarrow a_2 S S$

(GREIBACH NORMAL FORM, SIMPLE LL(1) GRAMMAR, LR(0) GRAMMAR)

EXAMPLE: $a_2 a_2 a_0 a_1 a_0 a_1 a_0$
DETERMINISTIC PUSHDOWN AUTOMATON FOR PREFIX NOTATION

OF TREE WITH NODES HAVING RANK 0, 1, 2:

\[
\begin{align*}
    a_0|S &\rightarrow \varepsilon & a_1|S &\rightarrow S \\
(1) &\quad S \rightarrow a_0 & \delta(0, a_0, S) &=(0, \varepsilon) \\
(2) &\quad S \rightarrow a_1 S & \delta(0, a_1, S) &=(0, S) \\
(3) &\quad S \rightarrow a_2 S S & \delta(0, a_2, S) &=(0, SS) \\
    a_2|S &\rightarrow SS \\
\end{align*}
\]

\[\delta(q, a, S) = \{(q, \alpha) : S \rightarrow a\alpha \in P\}\]
 NONDETERMINISTIC STRING SUFFIX AUTOMATON

EXAMPLE

![Diagram of a non-deterministic string suffix automaton]

(Vertices labeled 0 to 7 with transitions labeled 'a0', 'a1', and 'a2')
DETERMINISTIC STRING SUFFIX AUTOMATON
**SUBTREE PDA**

- **Ranked alphabet**
  \[ A = \{ a_2, a_1, a_0 \} \]

- **Tree**  \[ t_1 \]
  Prefix notation is
  \[ \text{pref}(t_1) = a_2 a_2 a_0 a_1 a_0 a_1 a_0 \]

- **Subtrees of**  \[ t_1 \] **in prefix notation are:**
  1.  \[ a_2 a_2 a_0 a_1 a_0 a_1 a_0 \]
  2.  \[ a_2 a_0 a_1 a_0 \]
  3.  \[ a_1 a_0 \]
  4.  \[ a_0 \]
All subtrees of tree $t_1$ and their prefix notation
Nondeterministic Subtree PDA $M_{nps}(t_1)$ for tree $t_1$ in prefix notation

$\text{pref}(t_1) = a2 \ a2 \ a0 \ a1 \ a0 \ a1 \ a0$
TRANSFORMATION TO DETERMINISTIC PDA

INPUT-DRIVEN PDA – pushdown store operations are determined by the input symbol.


MOREOVER, NONDETERMINISTIC ACYCLIC INPUT–DRIVEN PDA – THE CONTENTS OF THE PUSHDOWN STORE CAN BE PRECOMPUTED, AND ONLY TRANSITIONS AND STATES WITH POSSIBLE PUSHDOWN OPERATIONS ARE SELECTED.
Deterministic subtree PDA $M_{dps}(t_1)$ for tree in prefix notation $\text{pref}(t_1) = a2 a2 a0 a1 a0 a1 a0$
TRACE OF DETERMINISTIC SUBTREE PDA $M_{dps}(t_1)$ FOR AN INPUT SUBTREE $st$ IN PREFIX NOTATION $\text{pref}(st) = a_1a_0$

<table>
<thead>
<tr>
<th>STATE</th>
<th>PDS</th>
<th>INPUT</th>
<th>INPUT SUBTREE</th>
</tr>
</thead>
<tbody>
<tr>
<td>${0}$</td>
<td>$S$</td>
<td>$a_1\ a_0$</td>
<td>$a_1$</td>
</tr>
<tr>
<td>${4, 6}$</td>
<td>$S$</td>
<td>$a_0$</td>
<td></td>
</tr>
<tr>
<td>${5, 7}$</td>
<td>$\varepsilon$</td>
<td>$\varepsilon$</td>
<td>$a_0$</td>
</tr>
<tr>
<td>ACCEPT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Theorem 2**

Given a tree $t$ with $n$ nodes and its prefix notation $\text{pref}(t)$, the deterministic subtree PDA $M_{dps}(t)$ has just one pushdown symbol, fewer than $N \leq 2n + 1$ states and at most $N + n - 1 \leq 3n$ transitions.
WORK IN PROGRESS

EXTENSION FROM SUBTREES TO TREE PATTERNS:

JANOUŠEK, J., MELICHAR, B.: *Subtree Pushdown Automata and Tree Pattern Pushdown Automata for Trees in Prefix Notation*, UNPUBLISHED.
MORE INFORMATION ON

WEB PAGES http://www.arbology.org