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 $pref(t)$, ALL SUBTREES OF t IN PREFIX NOTATION ARE SUBSTRINGS OF $pref(t)$.

MOTIVATION

JANOUŠEK, J., MELICHAR, B.: *On Regular Tree Languages and Deterministic Pushdown Automata*, TO APPEAR IN ACTA INFORMATICA (SPRINGER), 2009

REGULAR TREE LANGUAGES ARE ACCEPTED BY **FINITE TREE AUTOMATA**.

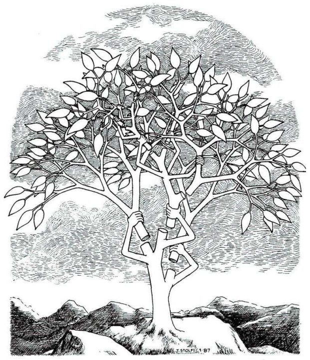
DETERMINISTIC PUSHDOWN AUTOMATA ACCEPT A PROPER SUPERCLASS OF THE REGULAR TREE LANGUAGES IN PREFIX OR POSTFIX NOTATION.

MOTIVATION

STRINGOLOGY
BASIC TOOL: FINITE AUTOMATA

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

ARBOLOGY AS AN ANALOGY WITH STRINGOLOGY



WEB PAGES http://www.arbology.org

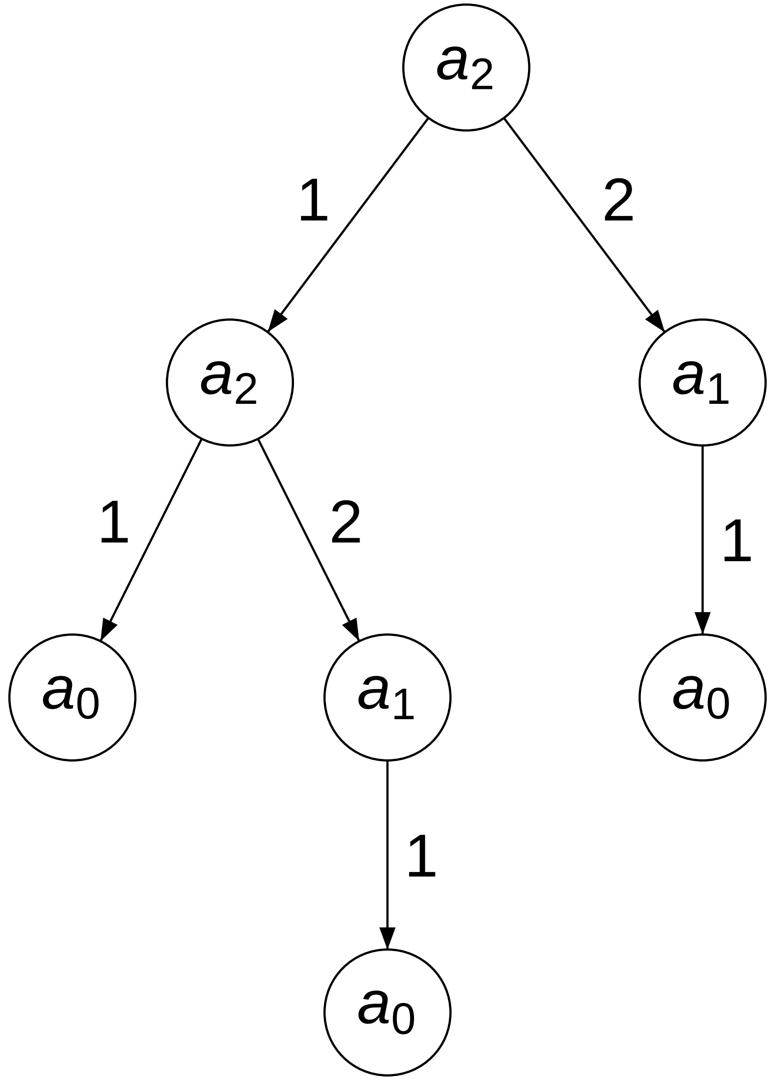
Prefix notation

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: $\underline{a_2\ \underline{a_2\ a_0\ \underline{a_1\ a_0}}\ \underline{a_1\ a_0}}$

6

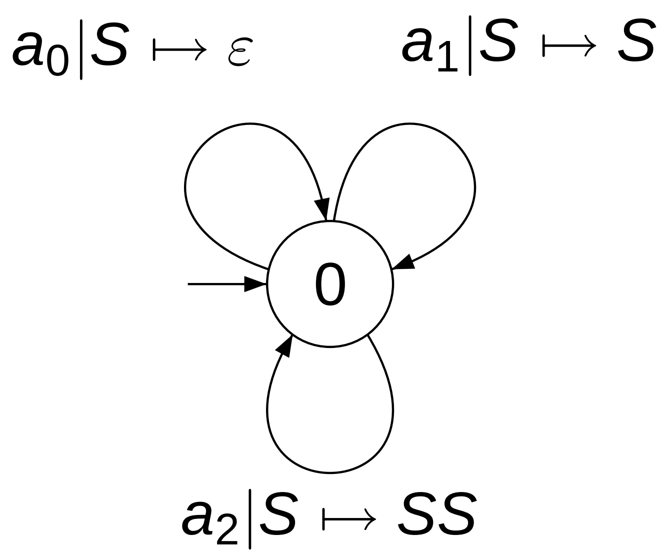
DETERMINISTIC PUSHDOWN AUTOMATON FOR PREFIX NOTATION

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

(1) $S \to a_0$

(2) $S \to a_1\ S$

(3) $S \to a_2\ S\ S$



$\delta(0, a_0, S) = (0, \varepsilon)$

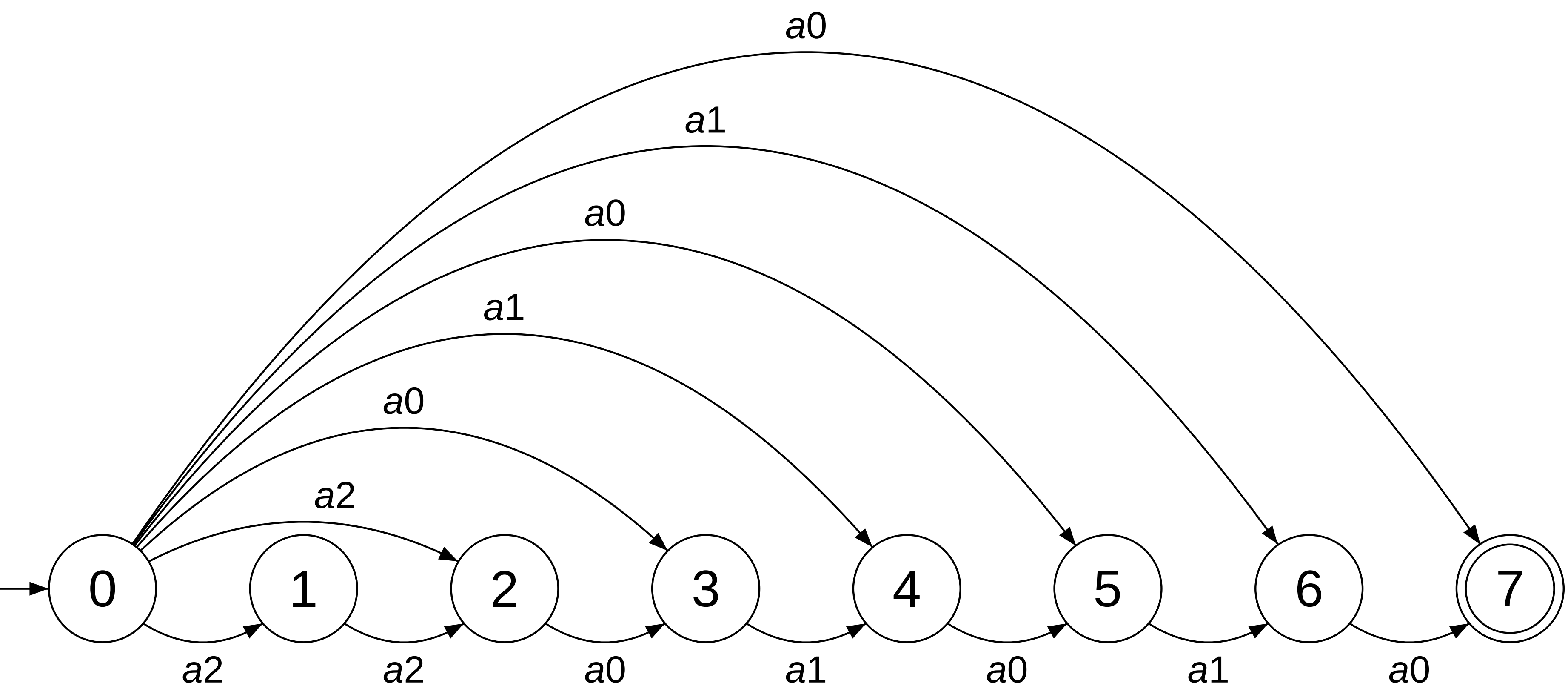
$\delta(0, a_1, S) = (0, S)$

$\delta(0, a_2, S) = (0, SS)$

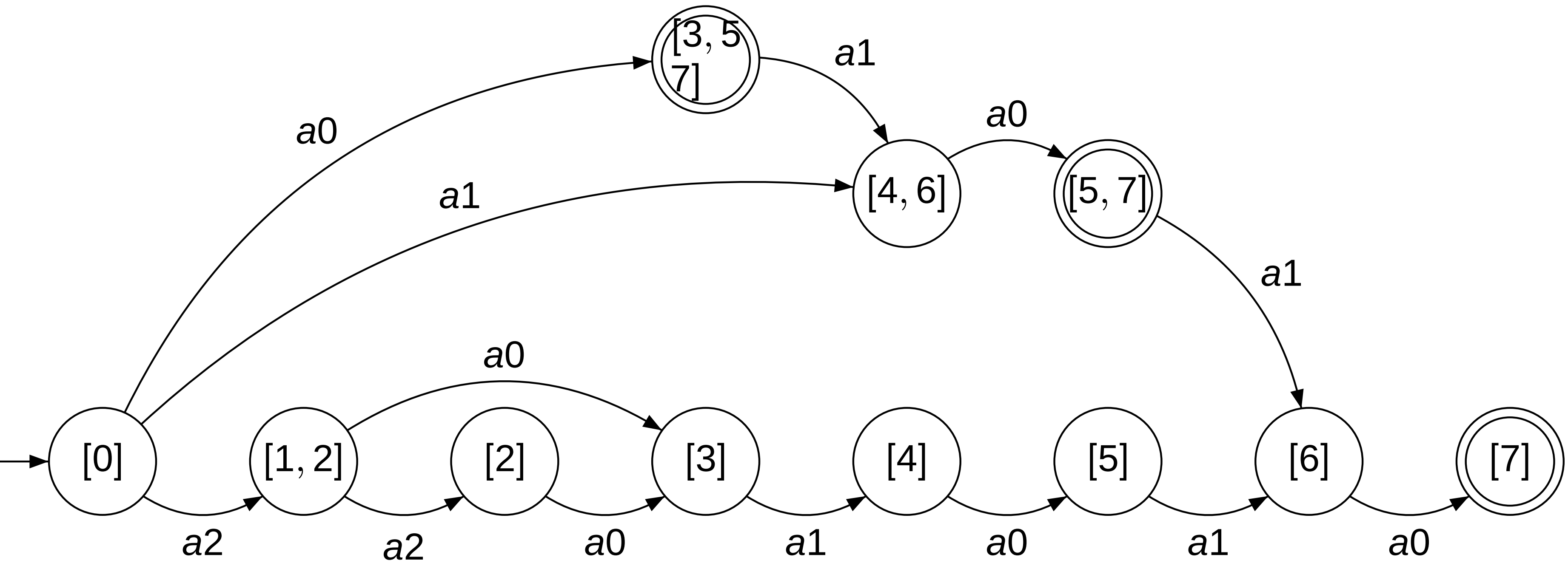
$$\delta(q, a, S) = \{(q, \alpha) : S \to a\alpha \in P\}$$

NONDETERMINISTIC STRING SUFFIX AUTOMATON

EXAMPLE



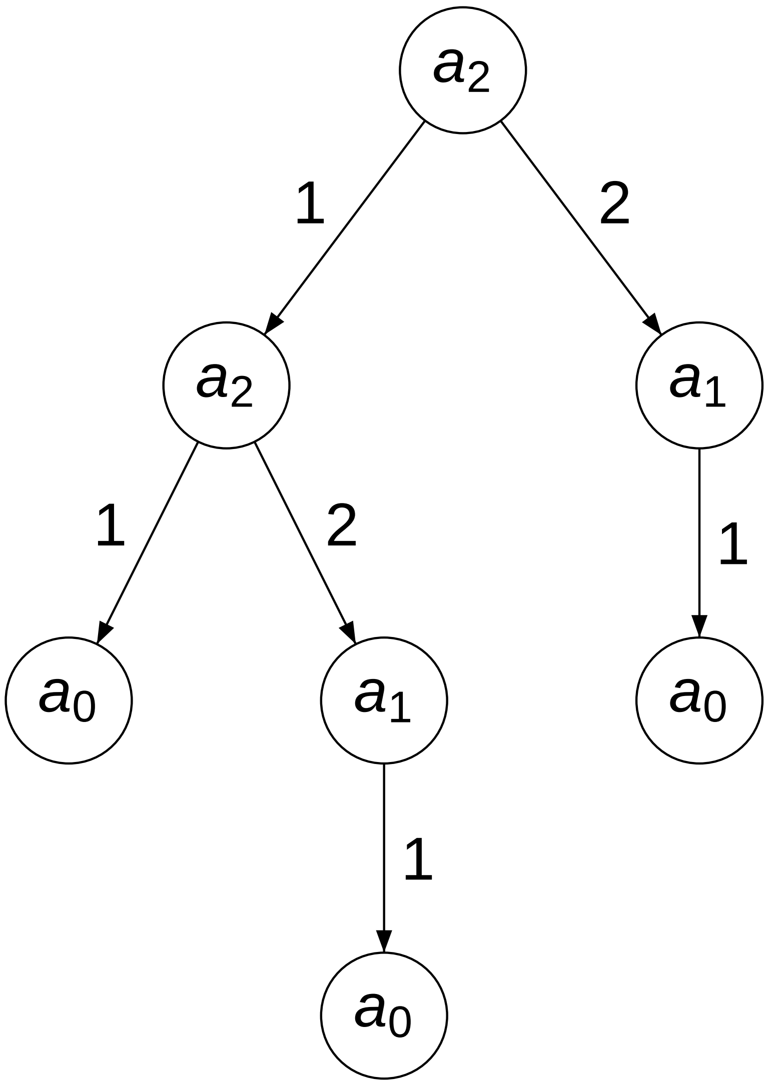
DETERMINISTIC STRING SUFFIX AUTOMATON



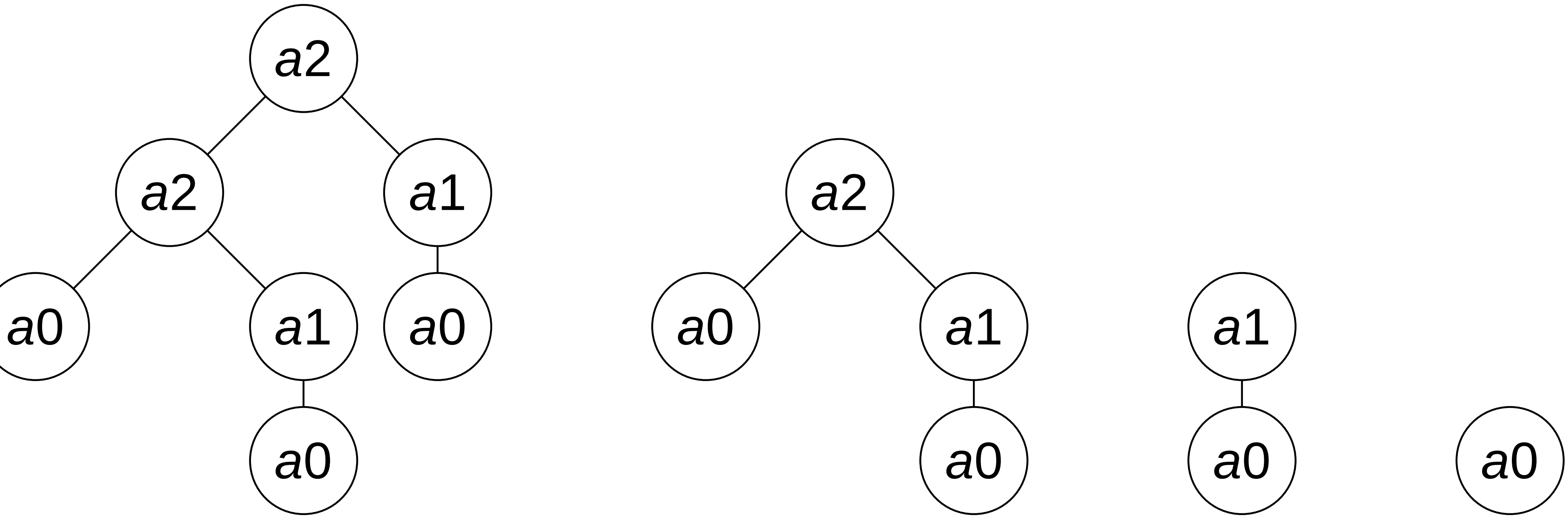
9

SUBTREE PDA

- RANKED ALPHABET $\mathcal{A} = \{a2, a1, a0\}$
- TREE t_1 PREFIX NOTATION IS $pref(t_1) =$ $a2\ a2\ a0\ a1\ a0\ a1\ a0$
- SUBTREES OF t_1 IN PREFIX NOTATION ARE:
  1. $a2\ a2\ a0\ a1\ a0\ a1\ a0$
  2. $a2\ a0\ a1\ a0$
  3. $a1\ a0$
  4. $a0$



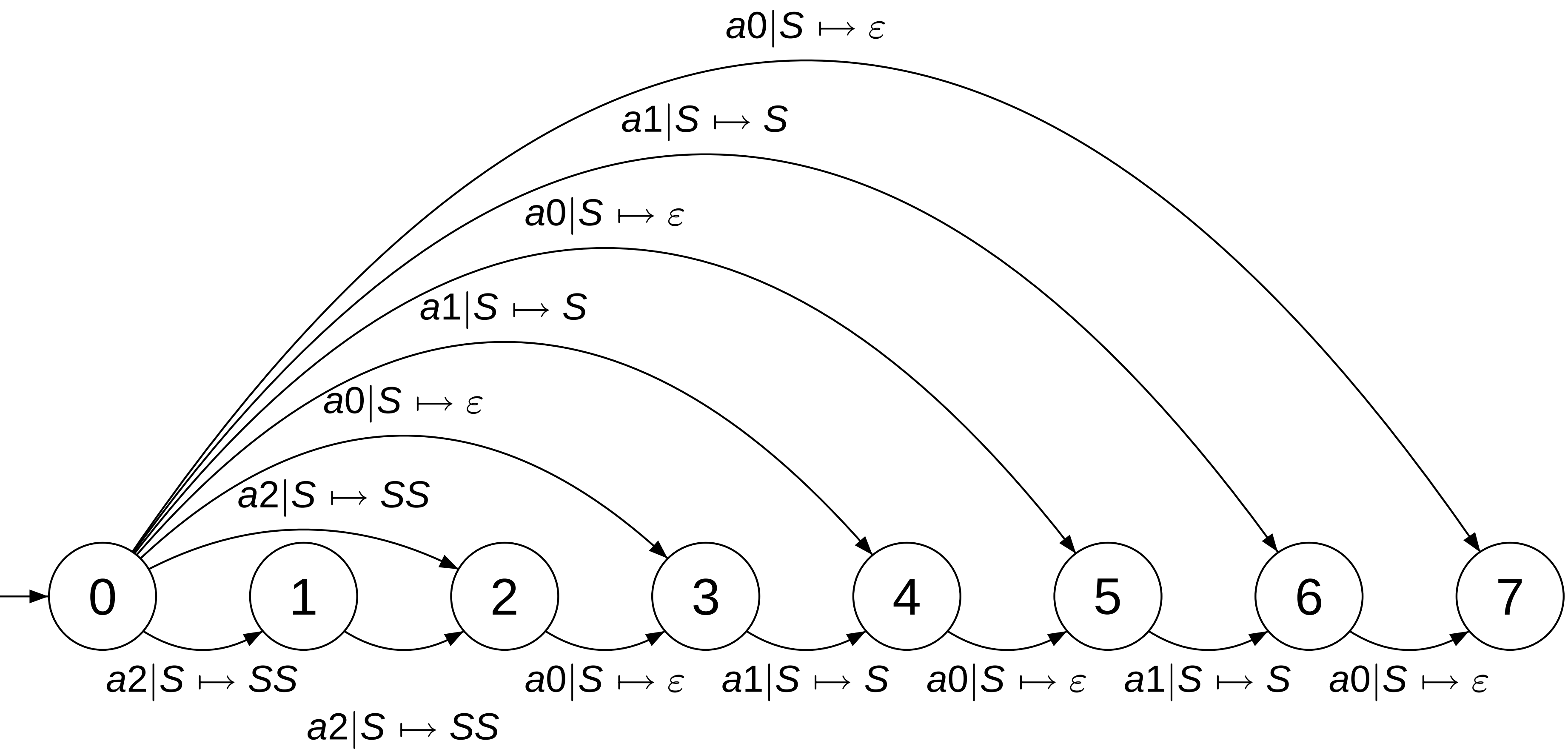
All subtrees of tree t_1 and their prefix notation



$a2\ a2\ a0\ a1\ a0\ a1\ a0$ $\quad$ $a2\ a0\ a1\ a0$ $\quad$ $a1\ a0$ $\quad$ $a0$

Nondeterministic subtree PDA $M_{nps}(t_1)$ for tree t_1 in prefix notation

$pref(t_1) = a2\ a2\ a0\ a1\ a0\ a1\ a0$



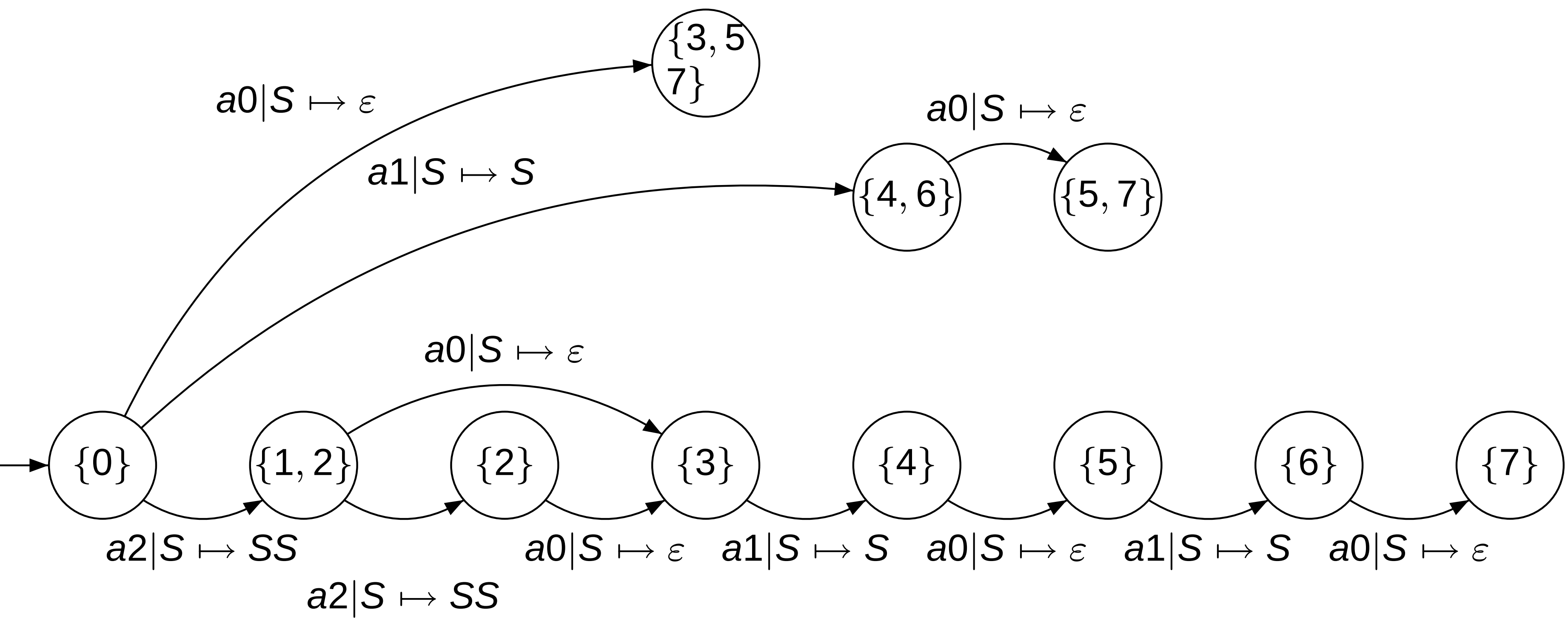
Transformation to Deterministic PDA

Input-driven PDA – pushdown store operations are determined by the input symbol.

Any nondeterministic input–driven PDA can be determinised similarly as in the case of finite automata – the states of the deterministic PDA correspond to subsets of states of the nondeterministic PDA (d–subsets).

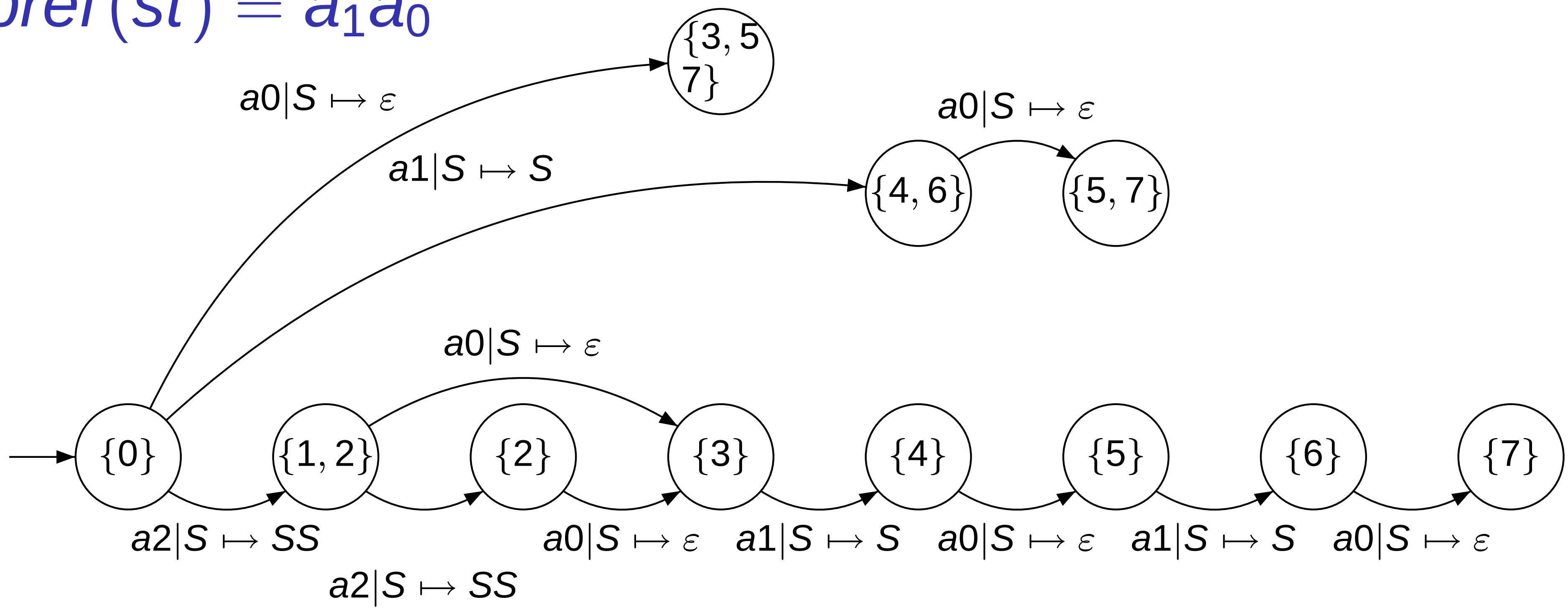
Moreover, nondeterministic acyclic input–driven PDA – the contents of the pushdown store can be precomputed, and only transitions and states with possible pushdown operations are selected.

DETERMINISTIC SUBTREE PDA $M_{dps}(t_1)$ FOR TREE IN PREFIX NOTATION $pref(t_1) = a2\ a2\ a0\ a1\ a0\ a1\ a0$



TRACE OF DETERMINISTIC SUBTREE PDA $M_{dps}(t_1)$ FOR AN INPUT SUBTREE *st* IN PREFIX NOTATION

$pref(st) = a_1 a_0$



STATE	PDS	INPUT
$\{0\}$	S	$a1\ a0$
$\{4, 6\}$	S	$a0$
$\{5, 7\}$	ε	ε
ACCEPT		

INPUT SUBTREE

$a1$
|
$a0$

15

THEOREM 2

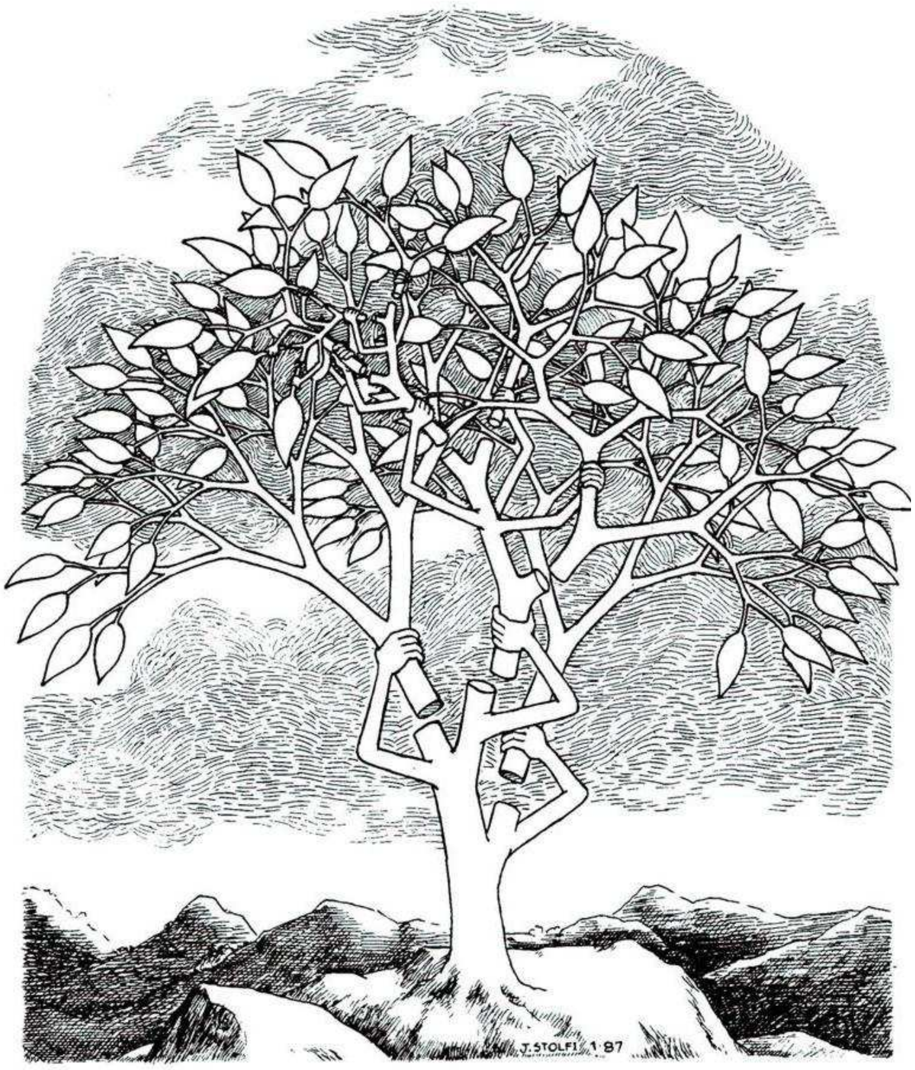
GIVEN A TREE t WITH n NODES AND ITS PREFIX NOTATION $pref(t)$, THE DETERMINISTIC SUBTREE PDA $M_{dps}(t)$ HAS JUST ONE PUSHDOWN SYMBOL, FEWER THAN $N \leq 2n + 1$ STATES AND AT MOST $N + n - 1 \leq 3n$ TRANSITIONS.

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`