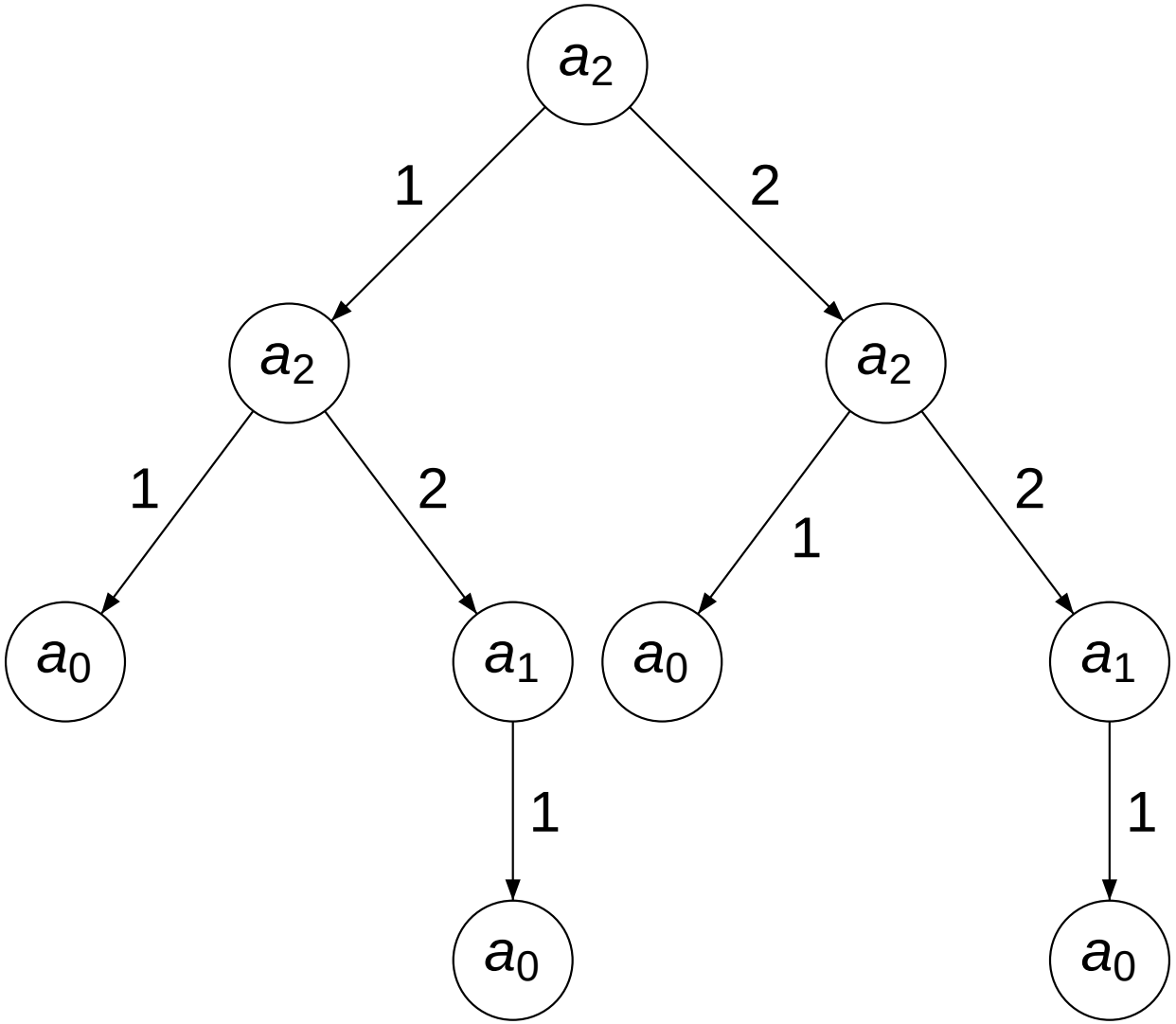
Repeats in Trees and Pushdown Automata

Bořivoj Melichar and Jan Janoušek

{melichar, janousej}@fel.cvut.cz

Department of Computer Science and Engineering
Faculty of Electrical Engineering
Czech Technical University
Czech Republic

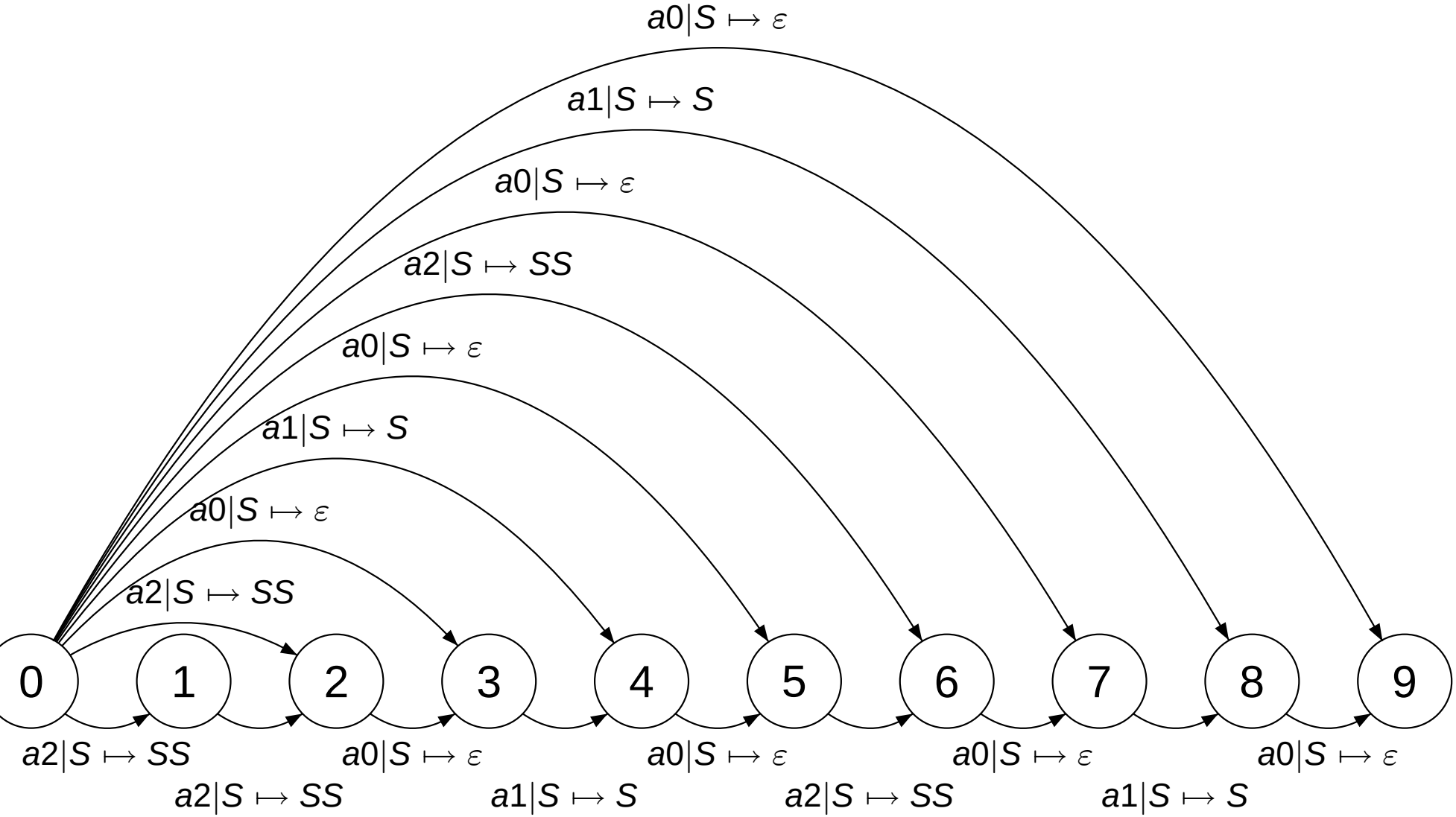
SUBTREE REPEATS



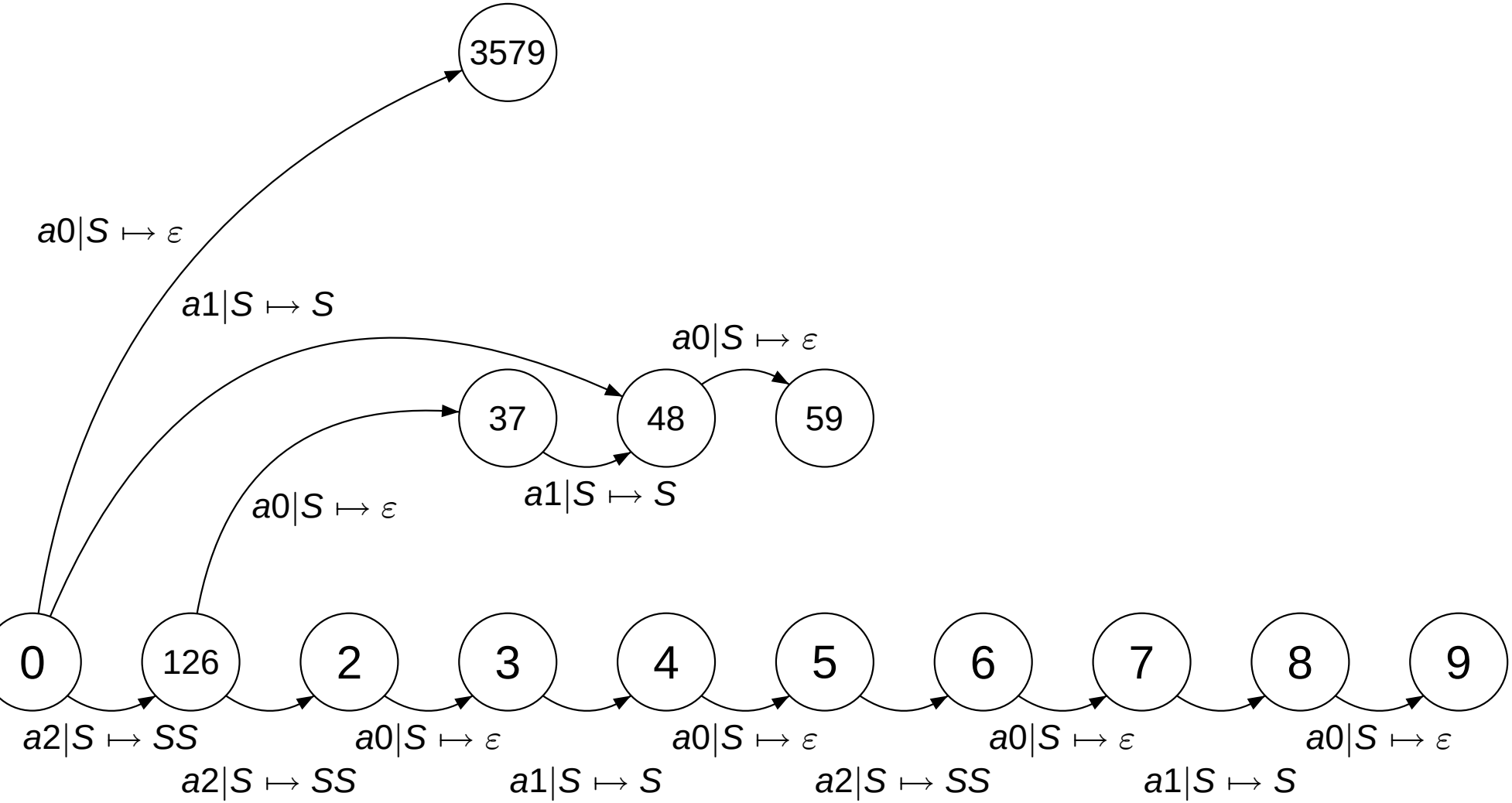
PREFIX NOTATION

*a*2 *a*2 *a*0 *a*1 *a*0 *a*2 *a*0 *a*1*a*0

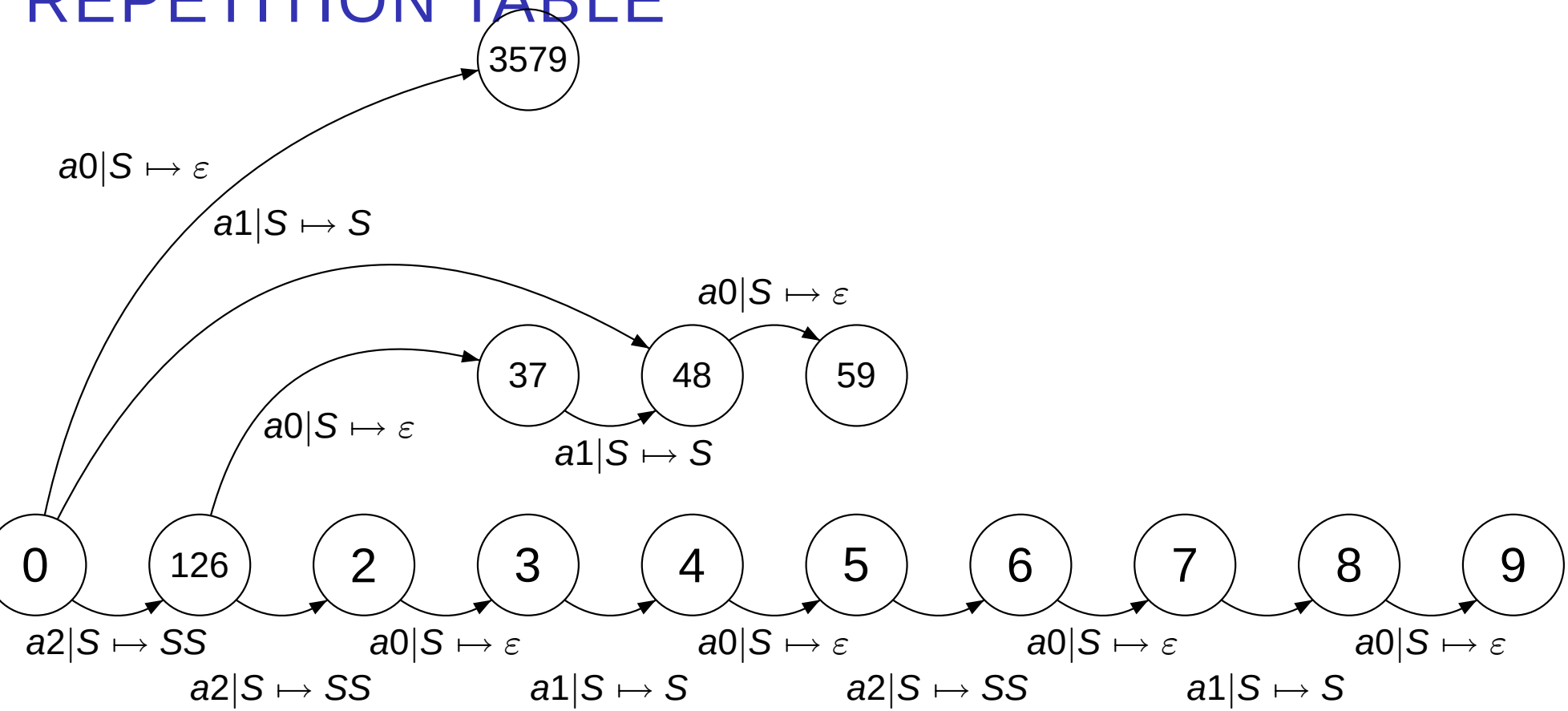
NONDETERMINISTIC SUBTREE PDA $M_{nps}(t_1)$ FOR $pref(t_1) = a2\ a2\ a0\ a1\ a0\ a2\ a0\ a1a0$



Deterministic subtree PDA $M_{dps}(t_1)$ for $pref(t_1) = a2\ a2\ a0\ a1\ a0\ a2\ a0\ a1a0$



REPETITION TABLE



D-SUBSET	SUBTREE	LIST OF REPEATS
3579	$a0$	$(3,F),(5,G),$
		$(7,G),(9,G)$
59	$a1\ a0$	$(5,F),(9,G)$
	$a2\ a0\ a1\ a0$	$(5,F),(9,N)$

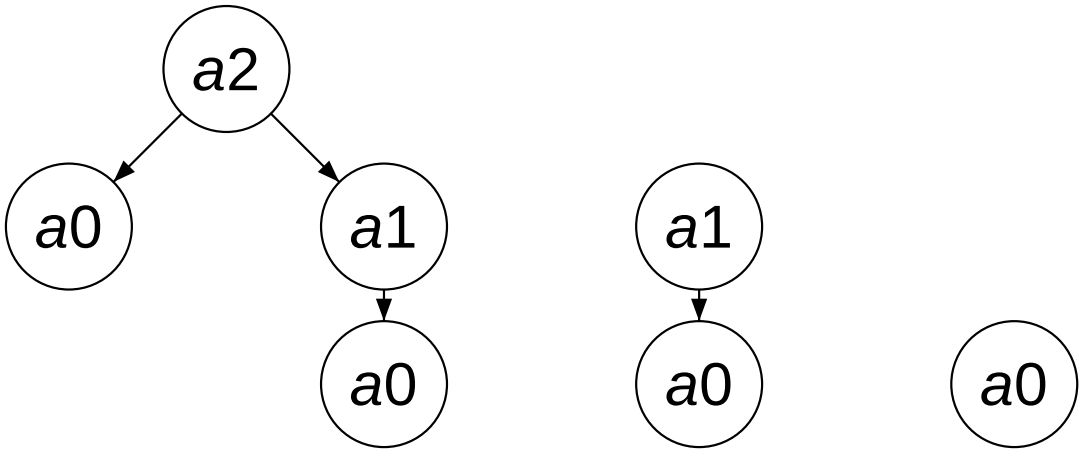
F FIRST

G GAP

N NEIGHBOUR

REPETITION TABLE

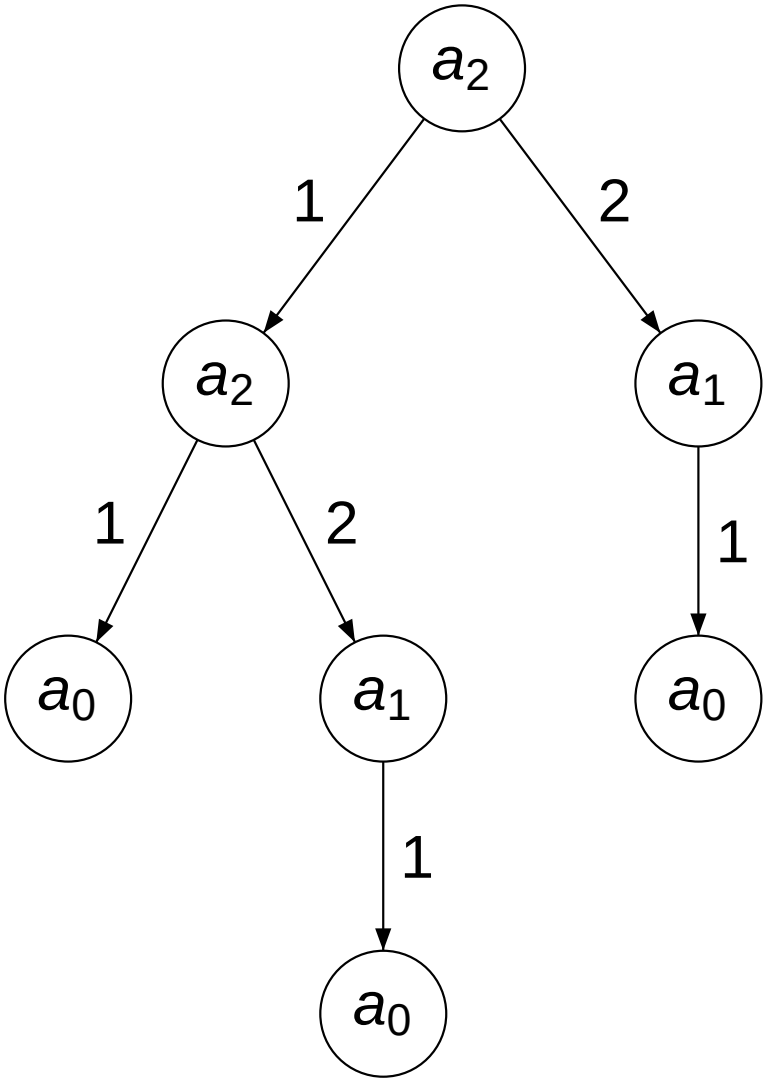
D-SUBSET	SUBTREE	LIST OF REPEATS
3579	$a0$	$(3,F),(5,G),(7,G),(9,G)$
59	$a1\ a0$	$(5,F),(9,G)$
	$a2\ a0\ a1\ a0$	$(5,F),(9,N)$



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

OVERLAPPING IS NOT POSSIBLE, WHICH FOLLOWS FROM THE BASIC PROPERTY OF TREE!

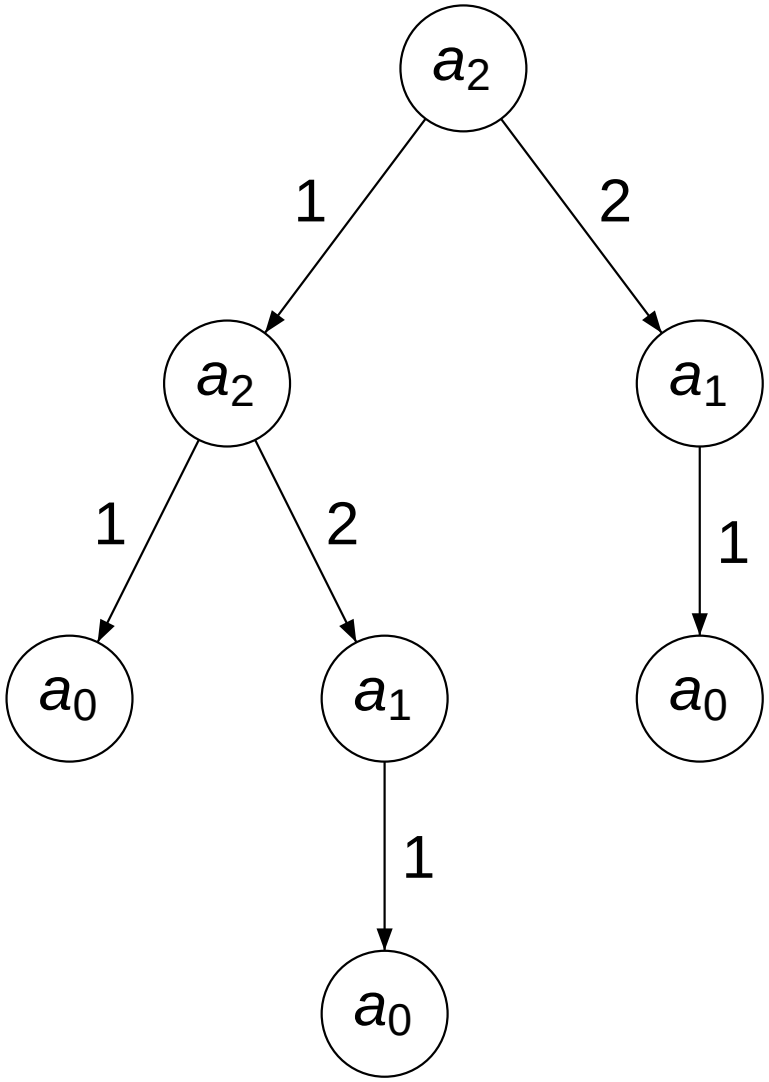
TREE PATTERNS REPEATS



PREFIX NOTATION

*a*2 *a*2 *a*0 *a*1 *a*0 *a*1 *a*0

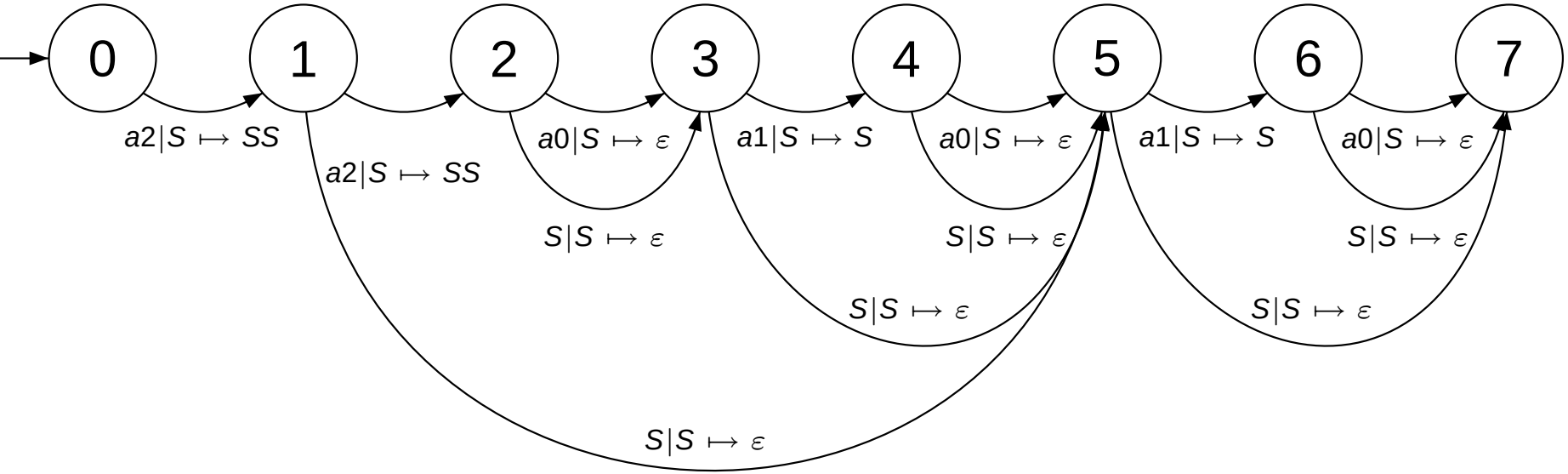
LIST OF SOME TOPS OF THE TREE



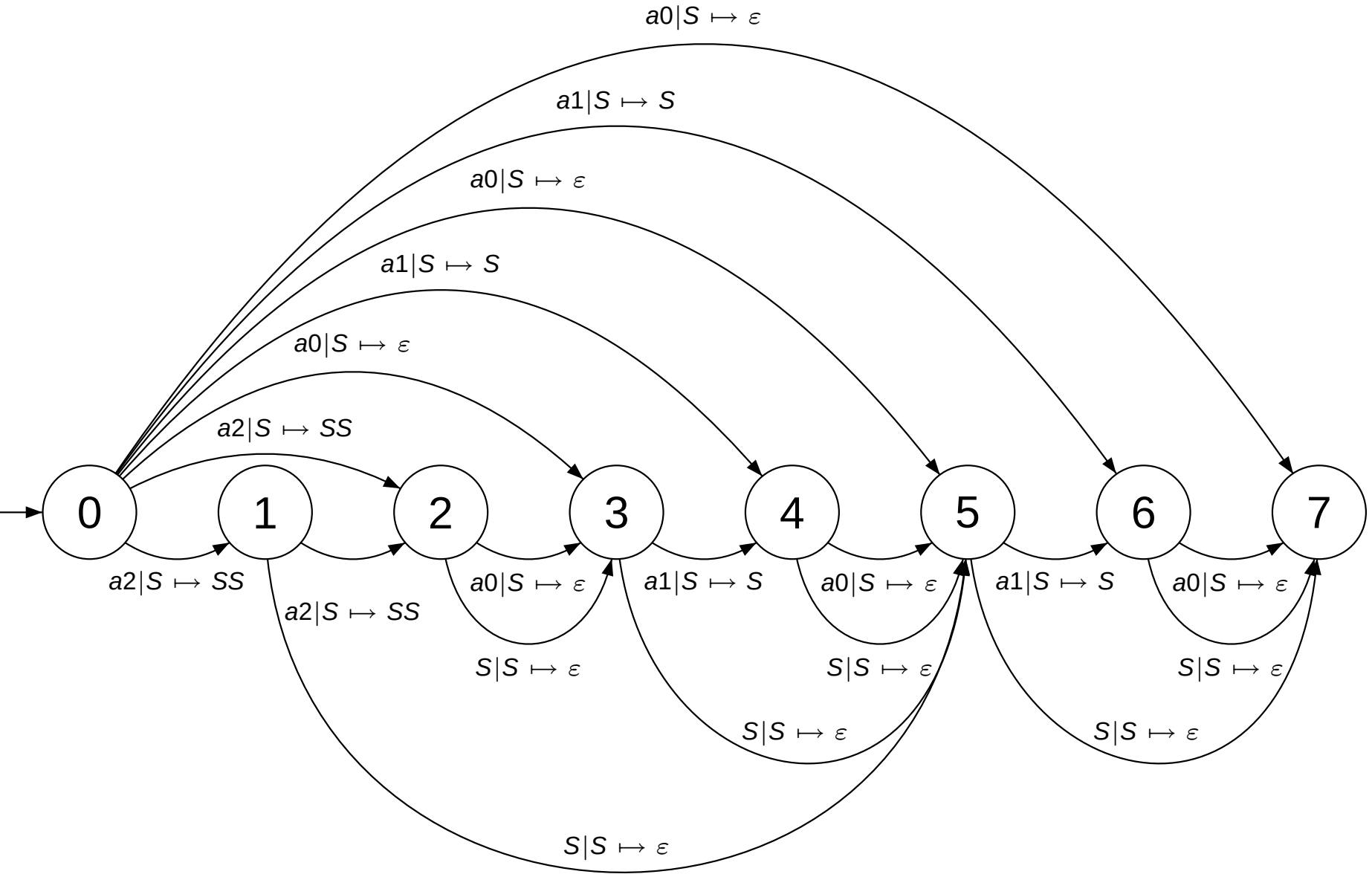
*a*2 *S S*
*a*2 *a*2 *S S S*
*a*2 *a*2 *a*0 *S S*
*a*2 *a*2 *a*0 *a*1 *S S*
*a*2 *a*2 *a*0 *a*1 *a*0 *S*
*a*2 *a*2 *a*0 *a*1 *a*0 *a*1 *S*
*a*2 *a*2 *a*0 *a*1 *a*0 *a*1 *a*0
*a*2 *S a*1 *S*
*a*2 *S a*1 *a*0
*a*2 *a*2 *S S a*1 *S*
.
.
.

Deterministic treetop PDA $M_{pt}(t_1)$ for $pref(t_1) = a2\ a2\ a0\ a1\ a0\ a1\ a0$

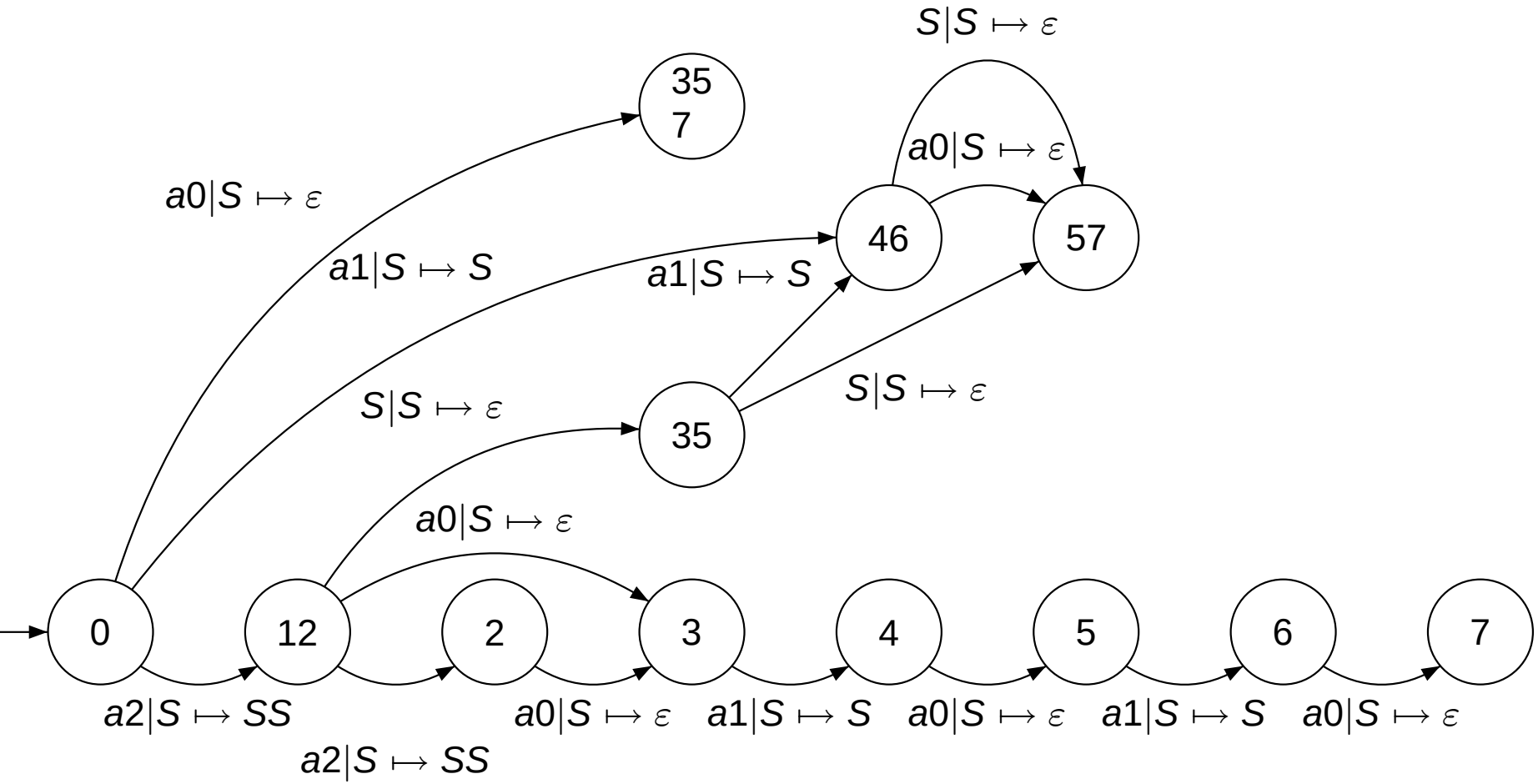
$srms = \{3, 5, 7\}$ Set of the Right-Most States



NONDETERMINISTIC TREE PATTERN PDA $M_{npg}(t_1)$ FOR $pref(t_1) = a2\ a2\ a0\ a1\ a0\ a1\ a0$



Deterministic tree pattern PDA $M_{dpg}(t_1)$ for $pref(t_1) = a2\ a2\ a0\ a1\ a0\ a1\ a0$



REPETITION TABLE

D-SUBSET	SUBTREE	LIST OF REPEATS
357	$a0$	$(3,F),(5,G),(7,G)$
57	$a1\ a0$	$(5,F),(7,G)$
	$a1\ S$	$(5,F),(7,G)$
	$a2\ SS$	$(5,F),(7,O)$
	$a2\ S\ a1\ S$	$(5,F),(7,O)$
	$a2\ S\ a1\ a0$	$(5,F),(7,O)$

F FIRST
G GAP
N NEIGHBOUR
O OVERLAPPING

COMPLEXITY

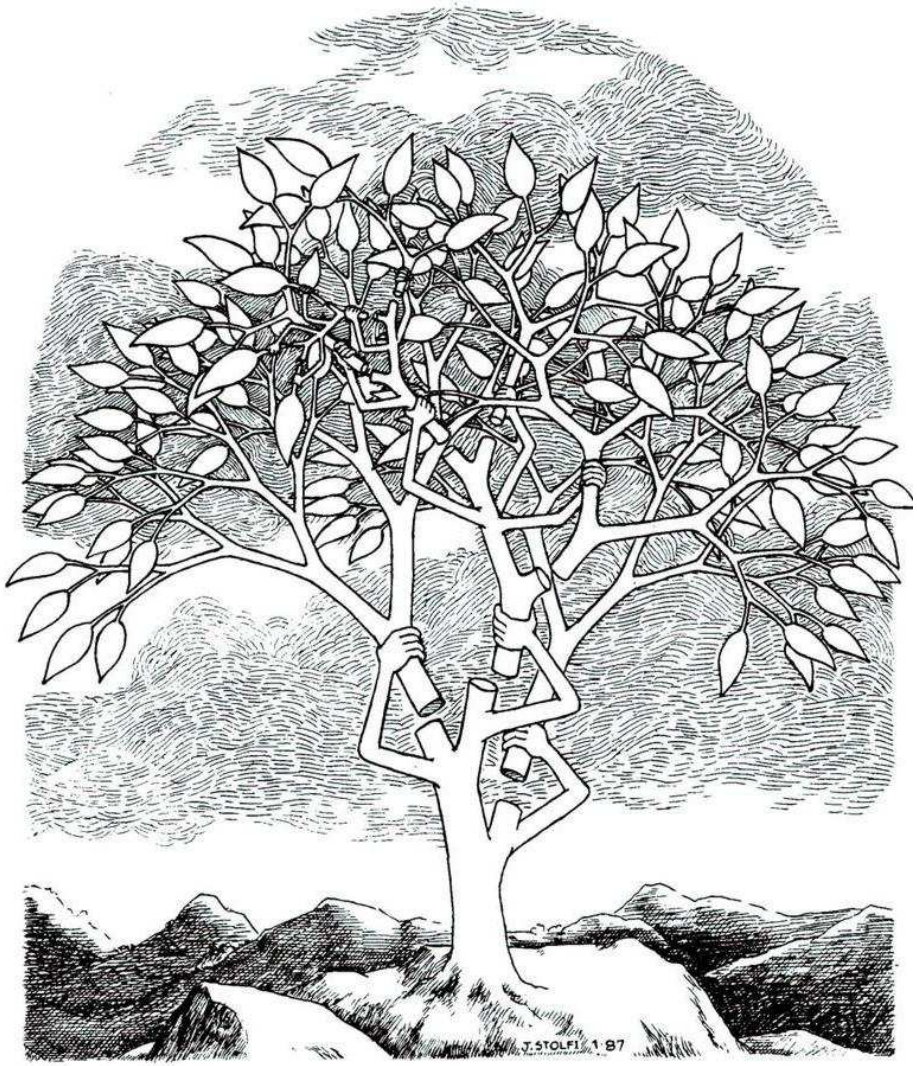
$$\mathcal{O}(n + r)$$

- n THE NUMBER OF NODES OF THE TREE
- r THE TOTAL SIZE OF REPEATING PARTS (SUBTREES, TEMPLATES) OF THE TREE (THE SIZE OF REPETITION TABLE)

$$r = \sum_p (rp * nr)$$

r IS THE TOTAL SIZE OF ALL PATHES FROM THE INITIAL STATE TO STATES WITH MULTIPLE SUBSETS.

- rp SIZE OF REPEATING PART
- nr NUMBER OF REPEATS (SIZE OF D–SUBSETS)
- p PATHES



WEB PAGES http://www.arbology.org

http://www.arbology.com

COMING SOON ...

Tree Languages, Tree Automata and Deterministic Pushdown Automata

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.

This is proved in:
Janoušek, J., Melichar, B.: *On Regular Tree Languages and Deterministic Pushdown Automata*. Submitted. 2008.

This paper contains also algorithm of transformation of any finite tree automaton to an equivalent deterministic pushdown automaton.