

Subtree Pushdown Automata for Trees in Bar Notation

JAN STOKLASA AND JAN JANOUŠEK AND BOŘIVOJ MELICHAR

{STOKLJA5, JANOUSEJ, MELICHAR}@FIT.CVUT.CZ

DEPARTMENT OF THEORETICAL COMPUTER SCIENCE
FACULTY OF INFORMATION TECHNOLOGY
CZECH TECHNICAL UNIVERSITY
CZECH REPUBLIC

BAR NOTATION

- A NEW LINEAR NOTATION FOR TREES
- ALLOWS TREE OPERATIONS CHANGING NODE ARITY ON LABELLED ORDERED TREES
- MORE GENERAL THAN THE BASIC PREFIX NOTATION OVER RANKED ALPHABET

MOTIVATION

STRINGOLOGY

String suffix and factor automata.

PROPERTIES:

- 1 ACCEPT ALL OCCURENCES OF AN INPUT SUFFIX AND AN INPUT FACTOR, RESPECTIVELY, IN A TEXT OF SIZE n .
- 2 SEARCH PHASE FOR ALL OCCURENCES OF AN INPUT SUFFIX OR AN INPUT FACTOR OF SIZE m IN TIME $\mathcal{O}(m)$, AND NOT DEPENDING ON n .
- 3 ALTHOUGH THE NUMBER OF FACTORS IN THE TEXT CAN BE QUADRATIC IN n , THE TOTAL SIZE OF THE DETERMINISTIC FACTOR AUTOMATON IS LINEAR IN n .

MOTIVATION

ARBOLOGY

Subtree and tree pattern pushdown automata – ANALOGOUS TO STRING SUFFIX AND FACTOR AUTOMATA.

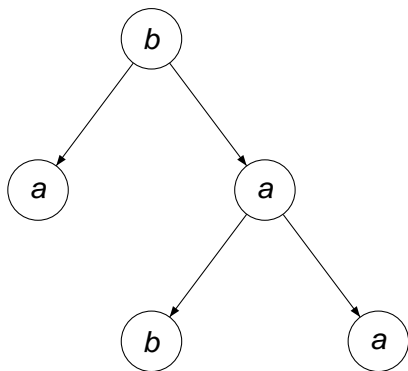
PROPERTIES:

- 1 ACCEPT ALL OCCURENCES OF AN INPUT SUBTREE IN A TREE OF SIZE n .
- 2 SEARCH PHASE FOR ALL OCCURENCES OF AN INPUT SUBTREE OF SIZE m IN TIME $\mathcal{O}(m)$, AND NOT DEPENDING ON n .

VARIABLE ARITY TREES

PREFIX BRACKETTED NOTATION

- ALPHABET $\mathcal{A} = \{a, b\}$
- TREE t_1
PREFIX BRACKETTED
NOTATION IS
 $pref_brac(t_1) =$
 $[b[a][a[b][a]]]$
- SUBTREES OF t_1 IN
PREFIX BRACKETTED
NOTATION ARE:
 - 1 $[b[a][a[b][a]]]$
 - 2 $[a[b][a]]$
 - 3 $[b]$
 - 4 $[a]$



THE LEFT-BRACKET IS REDUNDANT!

THE NEXT SYMBOL AFTER THE LEFT BRACKET IS ALWAYS THE ROOT
OF SUBTREE

BETTER SOLUTION: REMOVAL OF THE LEFT BRACKETS

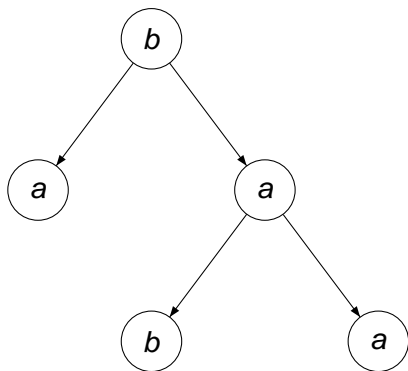
RESULT: PREFIX BAR NOTATION

VARIABLE ARITY TREES

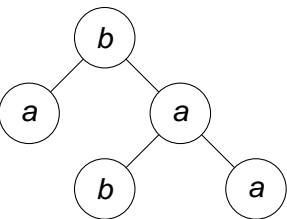
PREFIX BAR NOTATION

- ALPHABET $\mathcal{A} = \{a, b\}$
- TREE t_1
PREFIX BAR NOTATION
IS $pref_bar(t_1) =$
 $b a] a b] a]]]$
- SUBTREES OF t_1 IN
PREFIX BAR NOTATION
ARE:

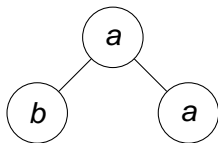
- 1 $b a] a b] a]]]$
- 2 $a b] a]]$
- 3 $b]$
- 4 $a]$



ALL SUBTREES OF TREE t_1 AND THEIR PREFIX BAR NOTATION



$b a] a b] a]]]$



$a b] a]]$



$b]$



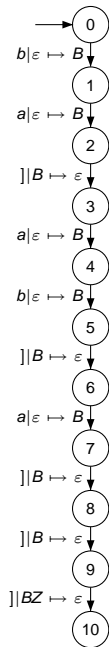
$a]$

THEOREM 1

GIVEN A TREE t AND ITS PREFIX BAR NOTATION $prefix_bar(t)$, ALL SUBTREES OF t IN PREFIX BAR NOTATION ARE SUBSTRINGS OF $prefix_bar(t)$.

EXAMPLE 1

TRANSITION DIAGRAM OF DETERMINISTIC PDA
 $M_p(t_1)$ ACCEPTING
 $prefix_bar(t_1) = b a] a b] a]]]$ BY EMPTY
PUSHDOWN STORE
INITIAL CONTENTS OF PUSHDOWN STORE IS Z



TRACE OF DETERMINISTIC PDA $M_p(t_1)$ FOR INPUT STRING $prefix_bar(t_1) = b a] a b] a]]]$

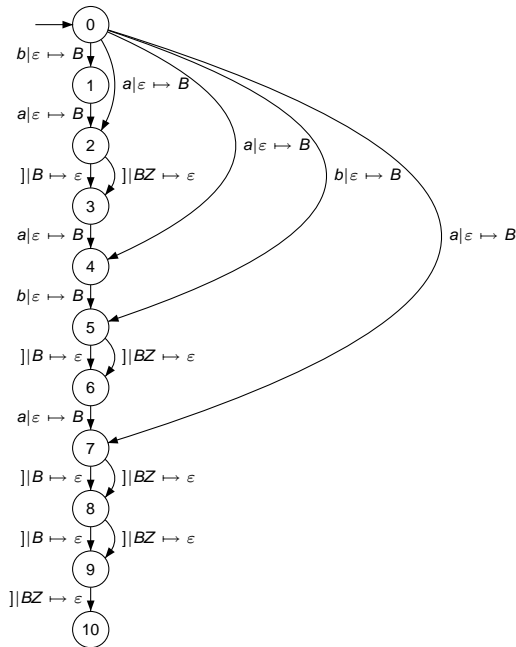
STATE	INPUT	PUSHDOWN STORE
0	$b a] a b] a]]]$	Z
1	$a] a b] a]]]$	BZ
2	$] a b] a]]]$	BBZ
3	$a b] a]]]$	BZ
4	$b] a]]]$	BBZ
5	$] a]]]$	BBBZ
6	$a]]]$	BBZ
7	$]]]$	BBBZ
8	$]]$	BBZ
9	$]$	BZ
10	ϵ	ϵ

ACCEPT

ACCEPT BY EMPTY PUSHDOWN STORE.

SUBTREE PDA

NONDETERMINISTIC
 SUBTREE PDA $M_{nps}(t_1)$
 FOR TREE t_1 IN PREFIX BAR
 NOTATION $prefix_bar(t_1) =$
 $b a] a b] a]]]$



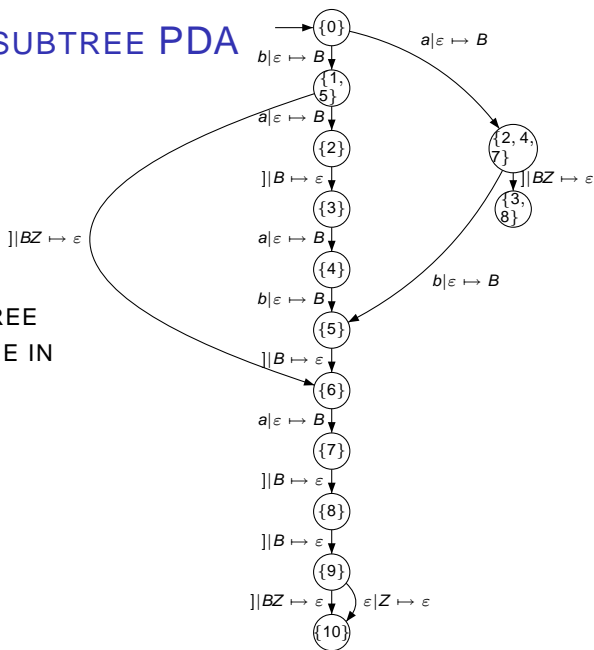
TRANSFORMATION TO DETERMINISTIC PDA

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



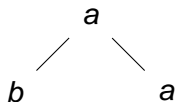
DETERMINISTIC SUBTREE
PDA $M_{dps}(t_1)$ FOR TREE IN
PREFIX BAR NOTATION

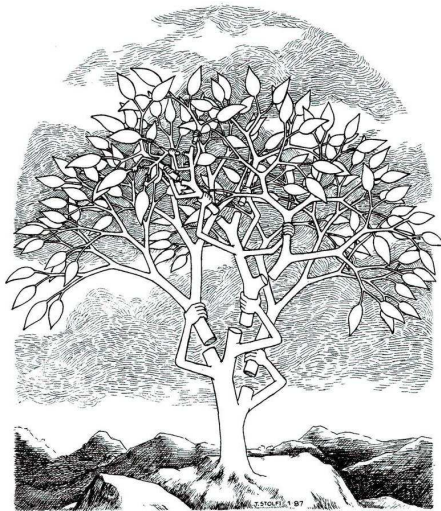
$prefix_bar(t_1) =$
 $b a] a b] a]]]$

TRACE OF DETERMINISTIC SUBTREE PDA $M_{dps}(t_1)$
 FOR AN INPUT SUBTREE st IN PREFIX BAR NOTATION
 $prefix_bar(st) = a b] a]]$

STATE	INPUT	PUSHDOWN STORE
{0}	$a b] a]]$	Z
{2, 4, 7}	$b] a]]$	BZ
{5}	$] a]]$	BBZ
{6}	$a]]$	BZ
{7}	$]]$	BBZ
{8}	$]$	BZ
{9}		Z
{10}		ϵ
ACCEPT		

INPUT SUBTREE





WEB PAGES <http://www.arbology.org>
<http://www.arbology.com>