

Januar 2015

RegexRangers

Naštevane besed regularnega jezika

Tadej Borovšak
Aleš Omerzel

Kazalo

1. Osnovne operacije
2. Drevesna struktura
3. Prevedba v avtomat
4. Obiskovanje avtomata
5. Capturing Groups
6. Problemi
7. Počitnice



Uvod: Kaj so regularni izrazi?

- Zaporedje znakov
- Vzorec za niz

Operacija	Primer
*	$ab^* = a, ab, abb, \dots$
+	$ab^+ = ab, abb, abbb, \dots$
?	$ab? = a, ab$
	$a b = a, b$
{n}	$ab\{3\} = abbb$
{n,}	$ab\{2, \} = abb, abbb, abbbb, \dots$
{n, m}	$ab\{2,3\} = abb, abbb$

Email: $/^{\wedge}([a-z0-9_ \backslash .-]^+)@([\backslash da-z \backslash .-]^+) \backslash .([a-z \backslash .]{2,6})\$/$

Naloga

- Naštevanje besed regularnega jezika
- Urejeno po
 - dolžini
 - abecedi
 - a, aa, ab, bbb, bcd, ...



1. Osnovne operacije

Izraz	Opis
a	Znak
e*	Kleene closure
e1e2	Catenation
e1 e2	Alternation

Primer dveh enostavnih izrazov:

- $a | ba^*c$
- $(a | b)^*c$



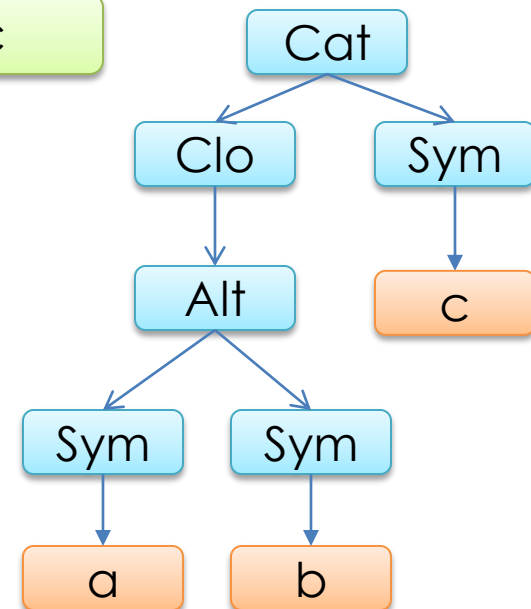
2. Drevesna struktura



- Prevedba regularnega izraza v drevesno strukturo

$(a | b)^*c \Rightarrow \text{Cat} (\text{Clo} (\text{Alt } a \ b)) \ c$

Izraz	Opis
\emptyset	Nil
ε	Eps
a	(Sym "a")
e^*	(Clo e)
$e_1 e_2$	(Cat e1 e2)
$e_1 e_2$	(Alt e1 e2)

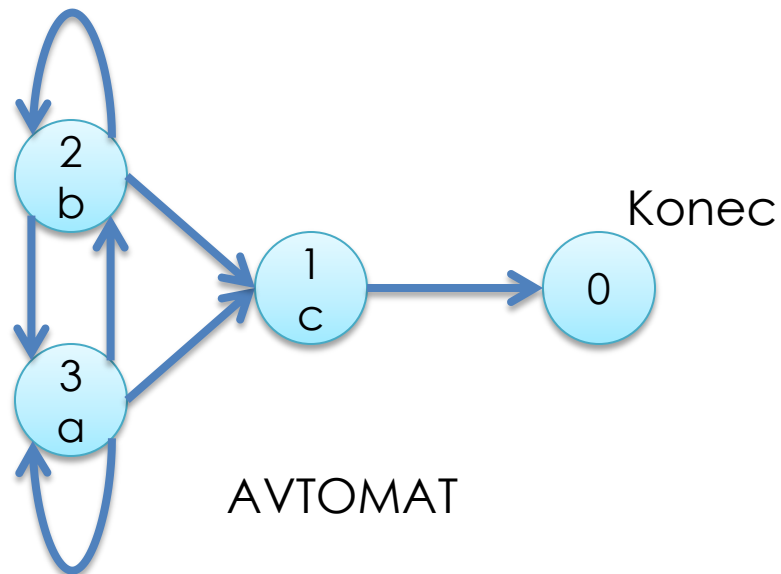


- Parser

3. Prevedba v avtomat



$(a | b)^* c$

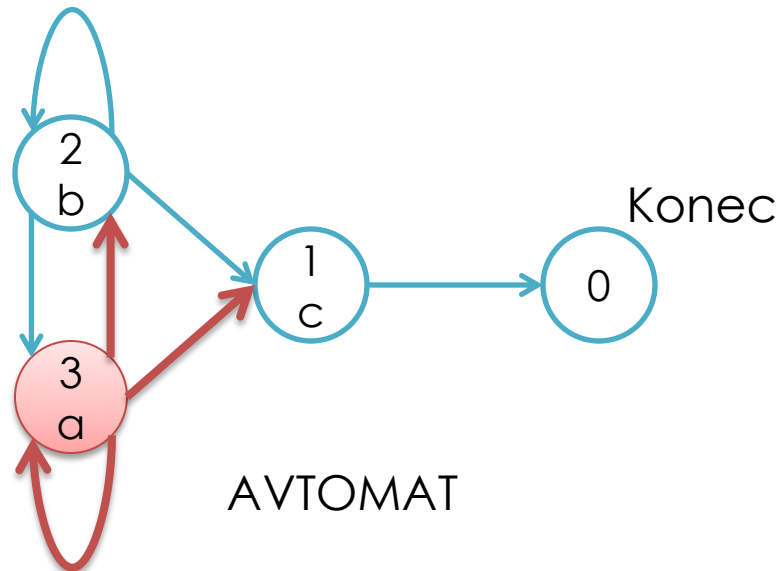


Kaj je avtomat? = graf

3. Prevedba v avtomat



$(a | b)^* c$



3. Prevedba v avtomat



$((abb)^* | ba)^* c | d^* | (ab)^*$

Prevedimo sedaj tole:

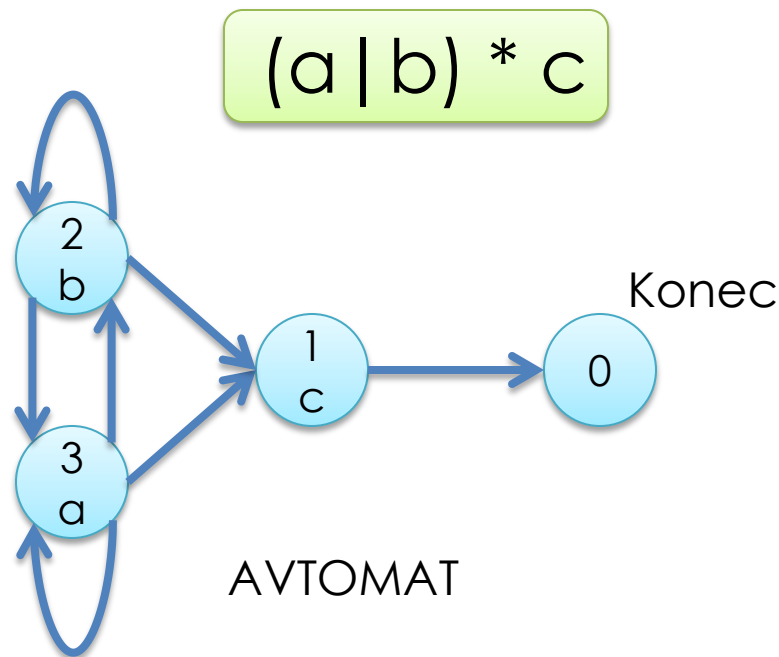
3. Prevedba v avtomat

$((abb)^* | ba)^* c | d^* | (ab)^*$



4. Obiskovalec avtomata

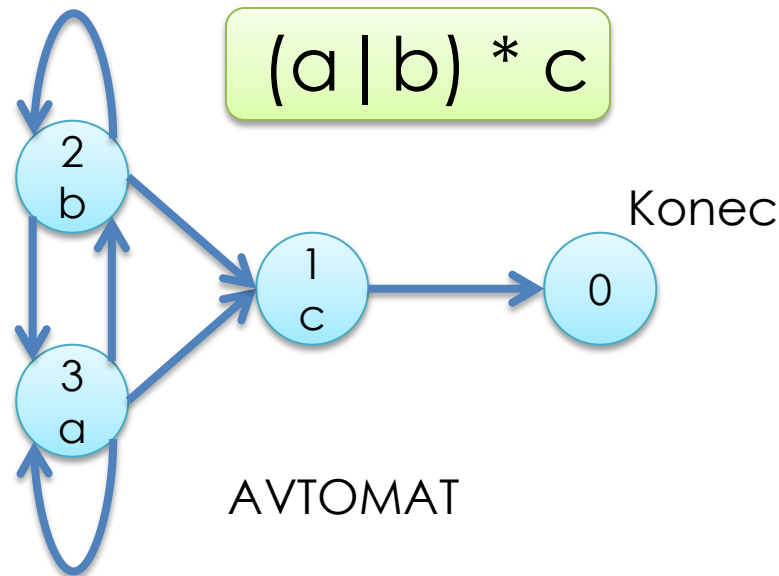
- = avanturist & printer
- Izpisuje besede urejene po **dolžini** in **abecedi** (CILJ?)



Izpis: [c, ac, bc, aac, abc, ...]

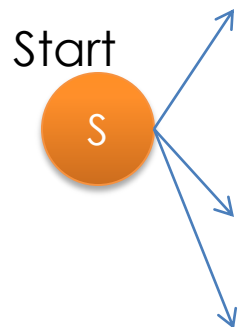


4. Obiskovalec avtomata

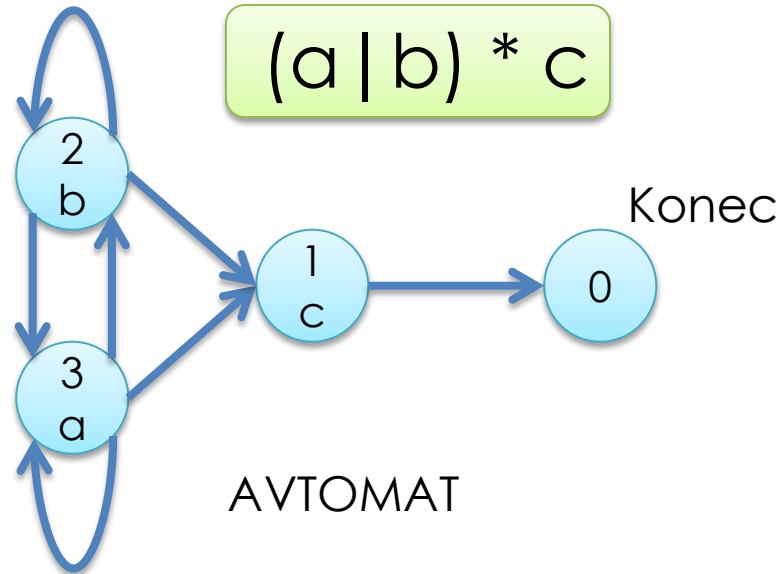


Vrsta stanj:

Izpis:

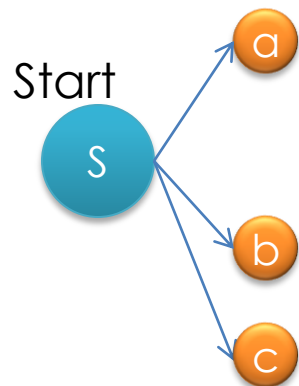


4. Obiskovalec avtomata

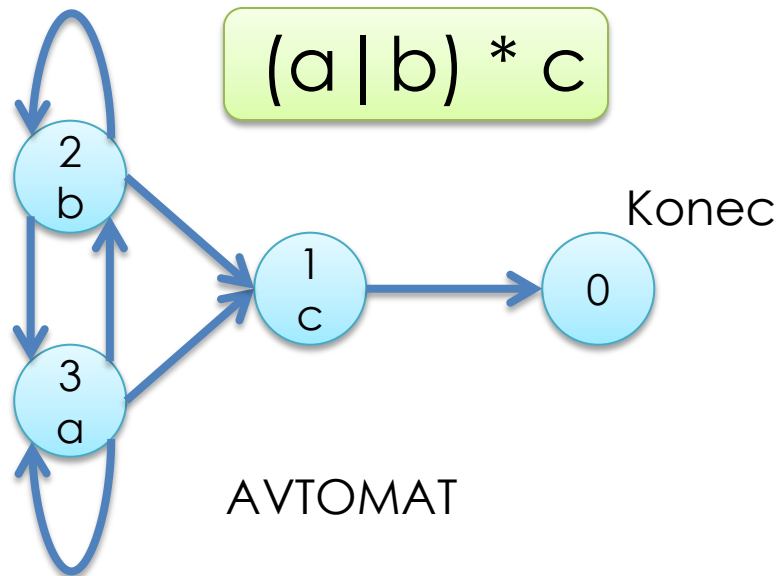


Vrsta stanj: + Stanje a, Stanje b, Stanje c

Izpis:

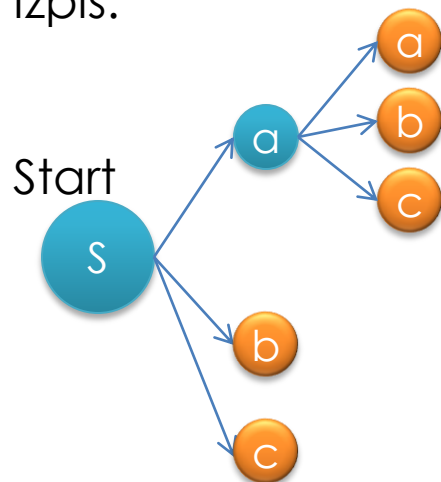


4. Obiskovalec avtomata

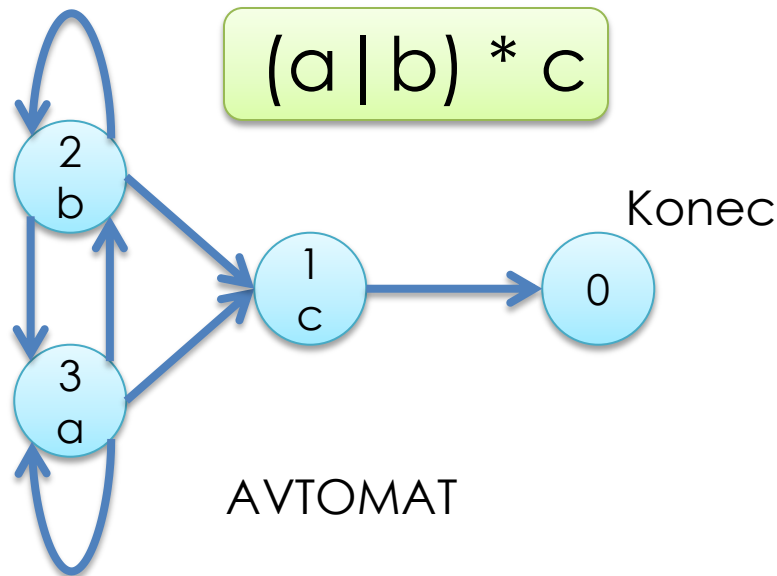


Vrsta stanj: ~~Stanje a, Stanje b, Stanje c,~~ + **Stanje a, Stanje b, Stanje c**

Izpis:

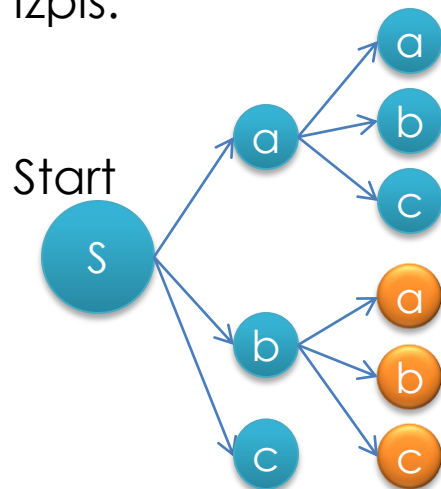


4. Obiskovalec avtomata

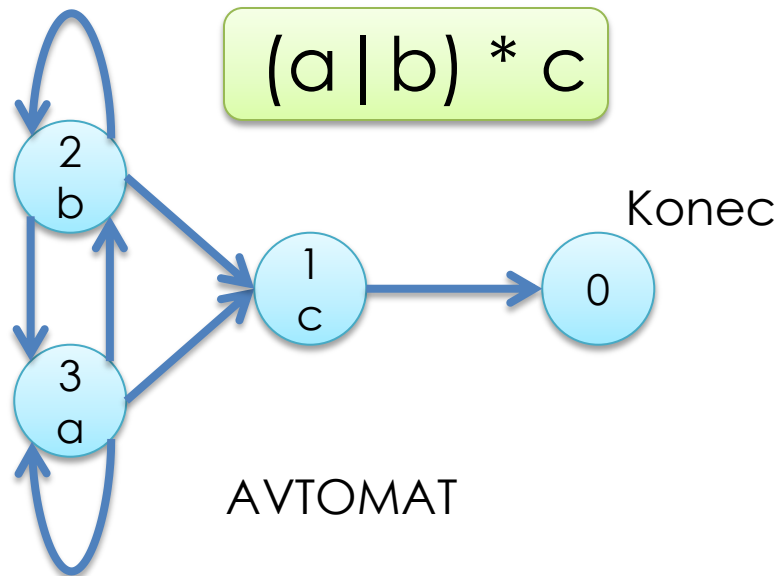


Vrsta stanj: ~~Stanje b, Stanje c, Stanje a, Stanje b, Stanje c,~~ + Stanje a, Stanje b, Stanje c

Izpis:

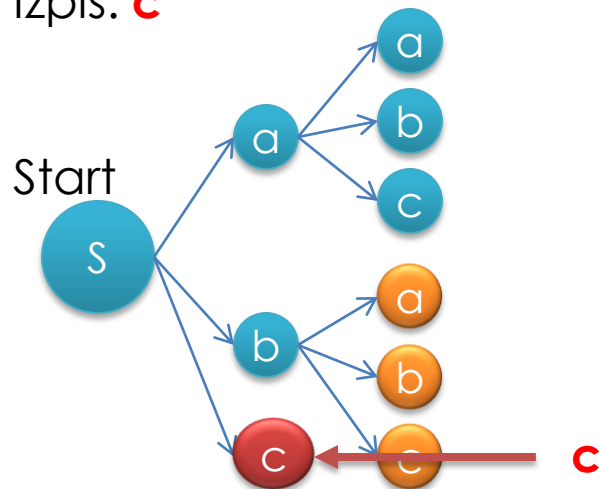


4. Obiskovalec avtomata

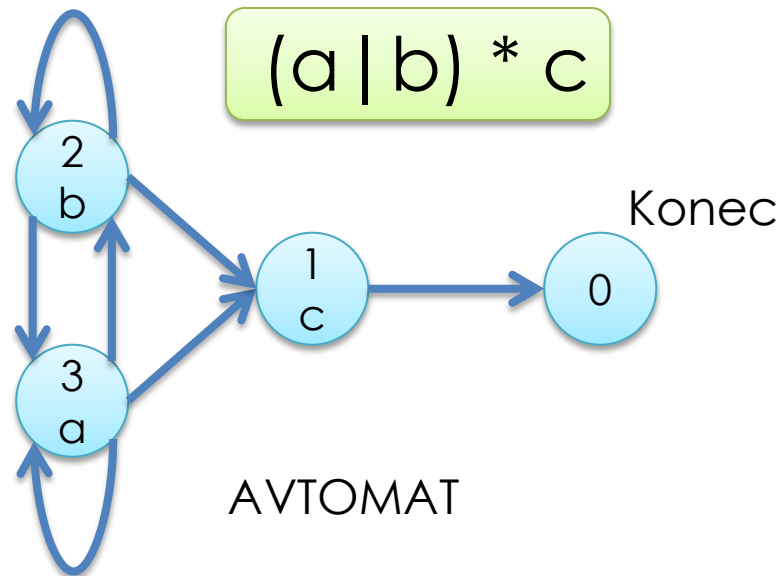


Vrsta stanj: ~~Stanje c, Stanje a, Stanje b, Stanje c~~, + **Stanje a, Stanje b, Stanje c**

Izpis: **c**

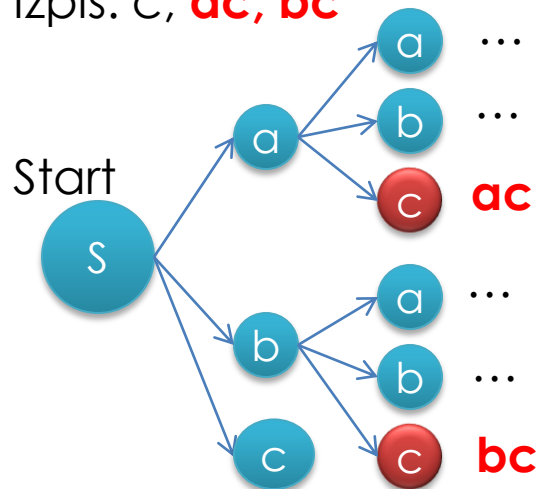


4. Obiskovalec avtomata



Vrsta stanj: Stanje a, Stanje b, Stanje c, + **Stanje a, Stanje b, Stanje c**

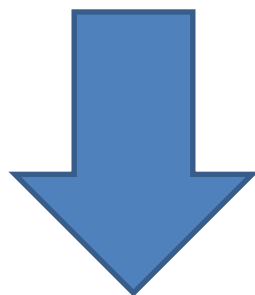
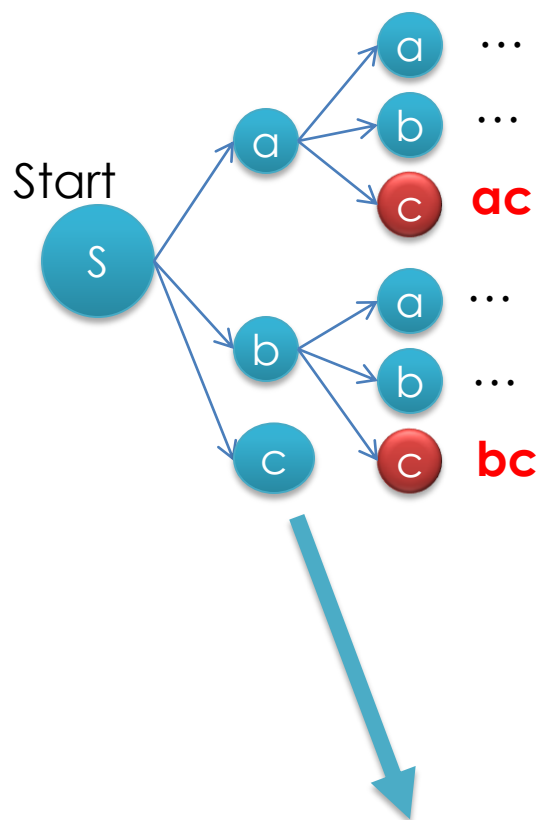
Izpis: c, **ac, bc**



Opomba

$$(a \mid b)^* c$$

Novo besedo dobimo, ko pridemo do **c**.


$$(\mathbf{c} \mid b)^* c$$

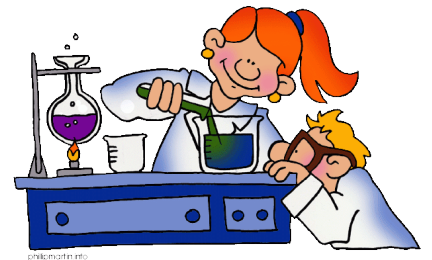
- Vsak c ima svoje stanje
- Indeksiranje

data State = State Ident Action NFA **deriving** Show



Demonstracija

- enumerate “ ab^* ”
- enumerate “ $a \mid b$ ”
- enumerate “ $(a \mid b)^*$ ”





Kaj še?

- Izpis dreves – po dolžini in abecedi
- Ne da zapisati z avtomatom
 - Lema o napihovanju
 - Ni kontekсно neodvisna gramatika
 - Zato ne obstaja avtomat





Capturing Groups

Capturing Groups???

$(ab) b (c) \mathbf{1} = abbcab$

$(ab) b (c) \mathbf{2} = abbcc$

- Kopiranje podniza
- Gnezdenje? (Memory)



Kako deluje?

- Obiskovalec vozlišč

- Špega po znakih

if “(“ then “start a new group”
if “)” then “close *the last* group”
if “2” then return group 2

NEW

- 1) memory za vse grupe
- 2) memory za trenutne grupe (gnezdenje)
 - za dodajanje znaka vsem trenutnim grupam

NEW



(a(b)) Memory grup: ab, b
Memory trenutnih: 1, 2 (če na poziciji b)

Problemi



- $a(b)^1*$ = ab **bbb**... ? ab **111**...
- $(a(b)^*)^2$ = a $b..b$ **b** ? a $b..b$ **b..b**
- $(a)(1)^2$ = a a **a** (run time) ? a a **1** (compile time)
- $(a | b^*)^1$ = ali je še urejen po dolžini?
- $(aaa | b^*)^1$ = ali je še urejen po dolžini?
- $a(bb)(1 | a)$ = ali je še urejen po dolžini? NE
- Kako pa zapisati številke v reg. izrazu? 5, 9, 14



Popravilo



1. $a(..)^+ \Rightarrow a \text{ **bc** **de** }$
 - $a(..)^+ (b)_2 \Rightarrow a(..)(..)_2(..) \dots b$
 - Množenje števila iste grupe = napaka
 - Potrebno prepisovanje iste grupe
2. Char \rightarrow data Action = Symbol Char
 - | Open
 - | Close
 - | Ref Int
 - | Accept
3. Popravilo parserja
4. Obiskovalec avtomata prilagoditi na Action

A tropical beach scene with a person relaxing in a wooden chair. The person is wearing a white straw hat and is seen from behind, sitting in a wooden lounge chair. The background features a clear blue sky, a turquoise ocean, and a green island in the distance. Palm fronds are visible in the top left corner. An orange banner is overlaid on the image, containing the word "Hollidays" in a white, stylized font.

Hollidays

Za konec še ...

- Testiranje
- Primeri
- Dokumentacija
- Program za konzolo



A tropical beach scene with a person relaxing in a wooden lounge chair. The person is wearing a white woven hat and is seen from behind. The chair is made of light-colored wood slats. The background features a clear blue sky, a turquoise ocean, and a green island in the distance. Palm fronds are visible in the top left corner. An orange banner with white text is overlaid on the image.

Hvala za pozornost