

Formalne metode u softverskom inženjerstvu

07 RE; Gramatike

ETFB 24-25

Dunja Vrbaški

$\text{RE} \rightarrow \text{e-NKA} \rightarrow \text{NKA} \rightarrow \text{DKA}$

$\text{RE} \leftarrow \text{e-NKA} \leftarrow \text{NKA} \leftarrow \text{DKA}$

$$\Sigma = \{0, 1\}$$

- Reči koje počinju znakom 0
-
- Reči koje imaju paran broj 0
- 0^*1110^*
- $\{0^n1^n \mid n \geq 1\}$

Koji jezici su regularni?

Da li je neki posmatrani jezik regularan ili ne?

Da li je proizvoljan konačan jezik regularan?

Šta je sa beskonačnim jezicima?

Ako postoji konačni automat (ili regularan izraz) - jeste regularan jezik.

T: Ako imamo konačan broj klasa ekvivalencije - jeste regularan jezik.

(relacija: x i y pripadaju istoj klasi ako za svaki nastavak z vazi xz je u jeziku akko yz je u jeziku; razlikovanje reči, ekvivalentne reči)

→ Ako postoji "konačan način ponašanja" stringova u jeziku - regularan jezik

T: Ako imamo konačan broj klasa ekvivalencije - jeste regularan jezik.

(relacija: x i y pripadaju istoj klasi ako za svaki nastavak z vazi xz je u jeziku akko yz je u jeziku; razlikovanje reči, ekvivalentne reči)

→ Ako postoji "konačan način ponašanja" stringova u jeziku - regularan jezik

Razmisliti: Koliko klasa ekvivalencije ima jezik $L = \{0^n1^n \mid n \geq 1\}$

Lema o napumpavanju za regularne jezike

Neka je L je beskonačan regularan jezik.

Tada postoji broj n takav da za sve reči $r \in L$ takve da $|r| \geq n$ postoji prefiks reči r koji sadrži pumpu.

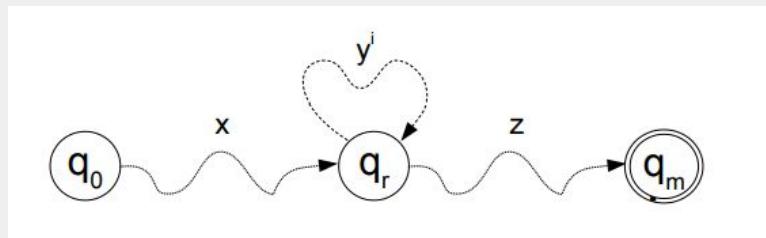
Odnosno, reč r se može zapisati u obliku:

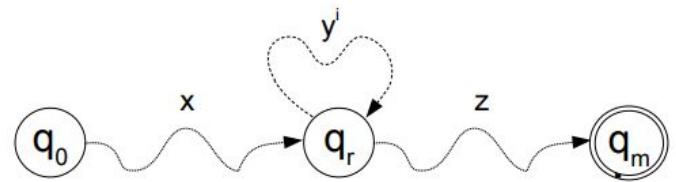
$$r=xyz$$

tako da

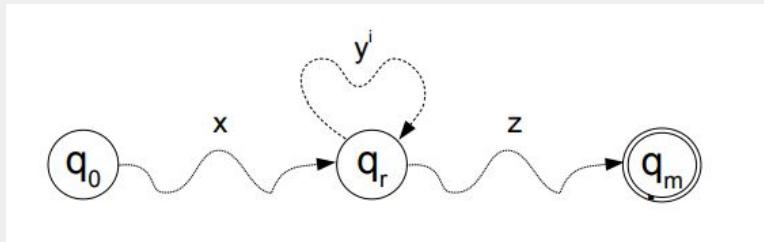
$$\begin{aligned}y &\neq \epsilon \\|xy| &\leq n \\xy^iz &\in L, i \geq 0.\end{aligned}$$

n je konstanta napumpavanja jezika L .





Svaki dovoljno dugačak string se može podeliti na tri dela gde se srednji deo može ukloniti ili ponavljati proizvoljan broj puta.



Obratiti pažnju:
pumpa - potreban, ali ne i dovoljan uslov

regularan jezik \rightarrow ima pumpu
ima pumpu \rightarrow ne mora da znači da je regularan

Može se koristiti za dokazivanje da jezik nije regularan.
(pp da je regularan, imamo pumpu, tražimo kontradikciju)

Zadatak

Dokazati da jezik $L = \{0^n1^n \mid n \geq 1\}$ nije regularan.

Idea:

PP da je regularan.

Imamo konstantu napumpavanja i xyz.

Tražimo reč koja dovodi do kontradikcije.

Nađimo neko i za koje xy^iz neće pripadati jeziku.

Zadatak

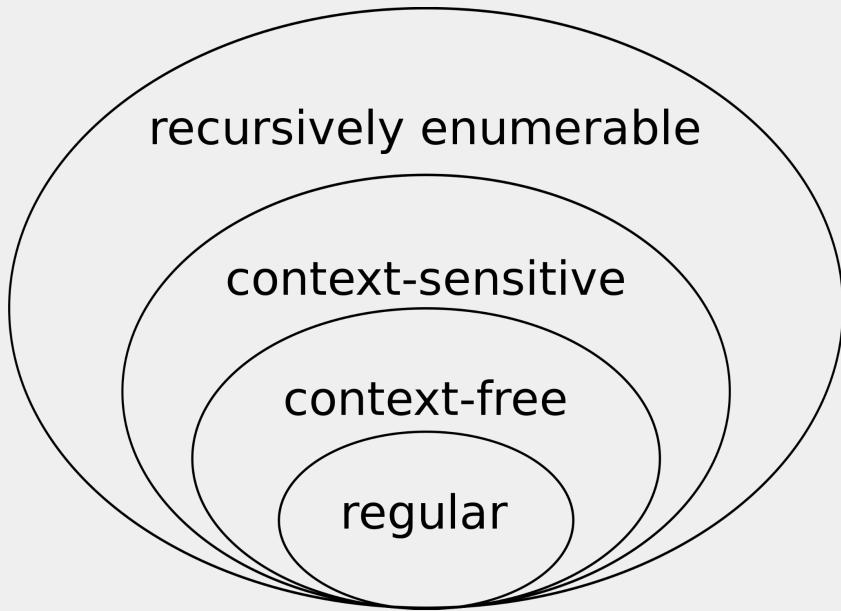
1.

Šta je Dirihićev (pigeonhole) princip?

2.

Razmisliti, ako već niste, o tome kako izgledaju automati (DKA, regularni jezici) i kako se odatile može doći do ideje pumping leme. Kakve veze ima broj stanja automata sa konstantom napumpavanja?

Hijerarhija Čomskog



Noam Chomsky

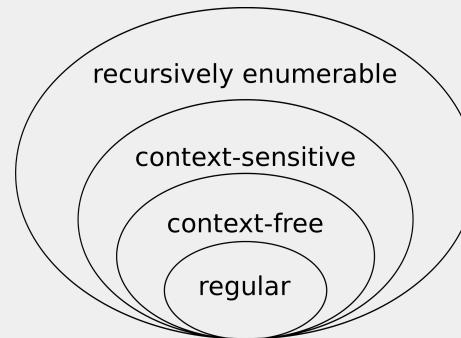
Inspiracija govorni jezici
(leksika, sintaksa, semantika)

→ formalni jezici

gramatike → generisanje
automati → prihvatanje

Za sada:
regularni izrazi → generisanje
DKA → prihvatanje

Nadskupovi:
moćnije, ekspresivnije
komplikovanije



Alfabet (azbuka) - konačan, neprazan skup simbola

$$\Sigma = \{0, 1\}$$

$$\Sigma = \{a, b, c\}$$

$$\Sigma = \{a, b, c, \dots z\}$$

$$\Sigma = \{ aa, ab, ac \} ?$$

Jezik - bilo koji podskup skupa Σ^*

$$\mathcal{L} \subseteq \Sigma^*$$

Kategorije?

Pravila po kojima se generišu reči?

Gramatika je uređena četvorka $G = (\Sigma, N, P, S)$ gde je:

- Σ skup simbola (terminala)
 - N skup pojmoveva (neterminala)
 - P skup pravila izvođenja
 - S početni pojam
-
- Σ, N, P su neprazni, konačni, skupovi
 - $\Sigma \cap N = \emptyset$
 - P je skup pravila u obliku $(\Sigma \cup N)^* N (\Sigma \cup N)^* \rightarrow (\Sigma \cup N)^*$

Gramatika je uređena četvorka $G = (\Sigma, N, P, S)$ gde je:

- Σ skup simbola (terminala)
- N skup pojmoveva (neterminala)
- P skup pravila izvođenja
- S početni pojam
- Σ, N, P su neprazni, konačni, skupovi
- $\Sigma \cap N = \emptyset$
- P je skup pravila u obliku $(\Sigma \cup N)^* N (\Sigma \cup N)^* \rightarrow (\Sigma \cup N)^*$

$$\begin{array}{lcl} S & \rightarrow & AB \\ A & \rightarrow & a \\ A & \rightarrow & aA \\ B & \rightarrow & b \\ B & \rightarrow & bB \end{array}$$

Jezik $L(G)$ generisan gramatikom G je skup reči nad skupom terminala koje se mogu izvesti počevši od početnog pojma.

Gramatika definiše pravila po kojima se grade reči (stringovi) u jeziku.

Svaka reč je morala nastati primenom pravila gramatike u određenom redosledu.

Izvođenje predstavlja korake koji su doveli do izgradenje reči.

$$\begin{array}{lcl} S & \rightarrow & AB \\ A & \rightarrow & a \\ A & \rightarrow & aA \\ B & \rightarrow & b \\ B & \rightarrow & bB \end{array}$$

$$L = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

Jezik je skup svih stringova koji se mogu izvesti iz početnog simbola S , primenom pravila gramatike.

\Rightarrow - jedno pravilo (jedan korak)

\Rightarrow^* - nula ili više koraka (ceo niz pravila)

Jezik $L = \{0^n 1^n \mid n \geq 1\}$

simboli: $\{0, 1\}$

pojmovi: $\{S\}$

početni pojam: S

pravila:

$$S \rightarrow 0S1$$

$$S \rightarrow 01$$

Primeri izvođenja

$$S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 000S111 \Rightarrow 00001111$$

$$S \Rightarrow^* 00001111$$

$$00001111 \in L$$

$$S \Rightarrow 01$$

$$S \Rightarrow^* 01$$

$$01 \in L$$

$$S \Rightarrow 0S1 \Rightarrow 0011$$

$$S \Rightarrow^* 0011$$

$$0011 \in L$$

$$\begin{array}{l} S \rightarrow aS \\ S \rightarrow bS \\ S \rightarrow \epsilon \end{array}$$
$$\begin{array}{l} A \rightarrow aA \\ A \rightarrow B \\ B \rightarrow bB \\ B \rightarrow \epsilon \end{array}$$
$$\begin{array}{l} S \rightarrow SS, \\ S \rightarrow (S), \\ S \rightarrow () \end{array}$$
$$\begin{array}{l} S \rightarrow aBC \\ S \rightarrow aSBC \end{array}$$
$$\begin{array}{l} CB \rightarrow CZ \\ CZ \rightarrow WZ \end{array}$$
$$\begin{array}{l} WZ \rightarrow WC \\ WC \rightarrow BC \end{array}$$
$$aB \rightarrow ab$$
$$bB \rightarrow bb$$
$$bC \rightarrow bc$$
$$cC \rightarrow cc$$
$$\begin{array}{l} S \rightarrow aSa, \\ S \rightarrow bSb, \\ S \rightarrow \epsilon, \end{array}$$
$$\begin{array}{l} S \rightarrow AX | XB \\ X \rightarrow aXb | \epsilon \\ A \rightarrow aA | a \\ B \rightarrow bB | b \end{array}$$
$$\begin{array}{l} S \rightarrow aSb \\ S \rightarrow ab \end{array}$$

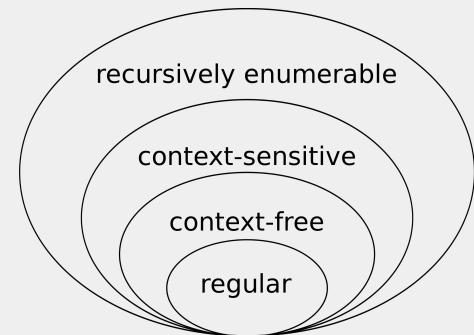
Notacija: oznaka za alternativu, umesto navođenja dva pravila ("dve strelice")

$$\begin{array}{l} S \rightarrow aA \\ A \rightarrow Sb \\ S \rightarrow \epsilon \end{array}$$

[Gramatika](#) je uređena četvorka $G = (\Sigma, N, P, S)$ gde je:

- Σ skup simbola (terminala)
 - N skup pojmoveva (neterminala)
 - P skup pravila izvođenja
 - S početni pojam
-
- Σ, N, P su neprazni, konačni, skupovi
 - $\Sigma \cap N = \emptyset$
 - P je skup pravila u obliku $(\Sigma \cup N)^* N (\Sigma \cup N)^* \rightarrow (\Sigma \cup N)^*$

Postavljanjem ograničenja na oblik pravila
→ klase gramatika
→ klase jezika



Regularna gramatika

type-3

P je skup pravila u obliku

$A \rightarrow S$

$A \rightarrow S B$

$A \rightarrow \epsilon$

ILI

$A \rightarrow S$

$A \rightarrow B S$

$A \rightarrow \epsilon$

$s \in \Sigma$

$A \rightarrow S$
 $A \rightarrow sB$
 $A \rightarrow \epsilon$

desno-regularne, desno-linearne

ILI

$A \rightarrow S$
 $A \rightarrow Bs$
 $A \rightarrow \epsilon$

levo-regularne, levo-linearne

$s \in \Sigma$

Linearne - najviše jedan neterminal sa desne strane

$S \rightarrow aA$
 $A \rightarrow Sb$
 $S \rightarrow \epsilon$

Linearna, ali nije regularna

Regularne gramatike
Regularan jezik
Regularni izrazi
Konačni automati

Desno-linearna gramatika G_1

$S \rightarrow abS$

$S \rightarrow a$

RE: $(ab)^*a$

Levo-linearna gramatika G_2

$S \rightarrow Aab$

$A \rightarrow Aab \mid B$

$B \rightarrow a$

RE: $a(ab)^*ab$

Zadatak

$S \rightarrow aS$

$S \rightarrow bA$

$A \rightarrow \epsilon$

$A \rightarrow cA$

Kako izgleda regularni izraz?

Zadatak

Da li se levo regularna gramatika može predstaviti kao desna i obrnuto?

Kontekstno slobodna gramatika

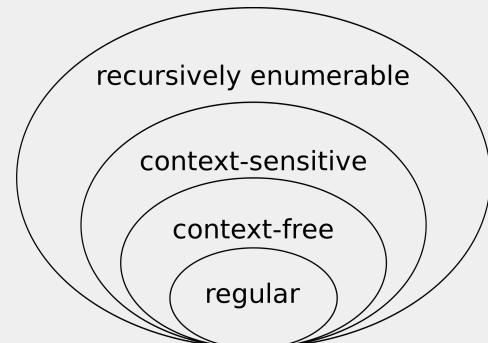
type-2

P je skup pravila u obliku

$A \rightarrow \alpha$

α - proizvoljan niz terminala i neterminala

I dalje imamo jedan neterminal sa leve strane, ali je desna strana opštija



$S \rightarrow aSa$

$S \rightarrow bSb$

$S \rightarrow \epsilon$

Koji je ovo jezik?

$$L = \{0^n 1^n \mid n \geq 1\}$$

Terminali: {0, 1}

Neterminali: {S}

Početni pojam: S

Pravila:

$$S \rightarrow 0S1$$

$$S \rightarrow 01$$

Nije regularan, nema DKA.

DKA ne ume da broji.

Da li postoje neki automati, moćniji od DKA, koji odgovaraju CFG jezicima?

Kontekstno zavisna (osetljiva) gramatika

type-1

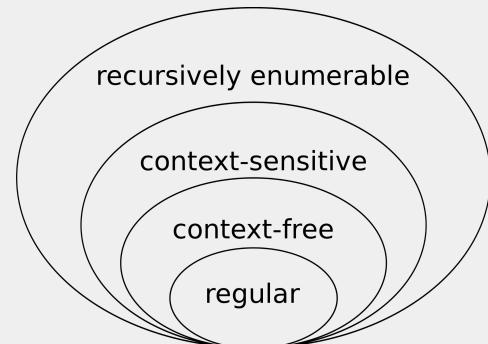
P je skup pravila u obliku

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

$$\alpha, \beta \in (\Sigma \cup N)^*$$

$$\gamma \in (\Sigma \cup N)^+$$

Više nije samo jedan neterminal sa leve strane.
Pojavljuje se kontekst.



$\{a^i b^i c^i \mid i \geq 1\}$

$S \rightarrow aBC$
 $S \rightarrow aSBC$

$CB \rightarrow CZ$
 $CZ \rightarrow WZ$
 $WZ \rightarrow WC$
 $WC \rightarrow BC$

$aB \rightarrow ab$
 $bB \rightarrow bb$
 $bC \rightarrow bc$
 $cC \rightarrow cc$

$\{a^i b^i c^i \mid i \geq 1\}$

1. $S \rightarrow aBC$
2. $S \rightarrow aSBC$
3. $CB \rightarrow CZ$
4. $CZ \rightarrow WZ$
5. $WZ \rightarrow WC$
6. $WC \rightarrow BC$
7. $aB \rightarrow ab$
8. $bB \rightarrow bb$
9. $bC \rightarrow bc$
10. $cC \rightarrow cc$

aabbcc

- $S \rightarrow_2 aSBC$
 $\rightarrow_1 aaB **$\rightarrow_3 aaBCZC$
 $\rightarrow_4 aaBWZC$
 $\rightarrow_5 aaBWCC$
 $\rightarrow_6 aaBBCC$
 $\rightarrow_7 aabBCC$
 $\rightarrow_8 aabbCC$
 $\rightarrow_9 aabbC$
 $\rightarrow_{10} aabbcc$**$

Zadatak

Prikazati izvođenje za aaabbbccc.

$$\begin{array}{l} S \rightarrow aSa \\ S \rightarrow bSb \\ S \rightarrow \epsilon \end{array}$$
$$\begin{array}{l} S \rightarrow aBC \\ S \rightarrow aSBC \end{array}$$
$$\begin{array}{l} CB \rightarrow CZ \\ CZ \rightarrow WZ \\ WZ \rightarrow WC \\ WC \rightarrow BC \end{array}$$
$$\begin{array}{l} aB \rightarrow ab \\ bB \rightarrow bb \\ bC \rightarrow bc \\ cC \rightarrow cc \end{array}$$

kontekstno slobodna

kontekstno zavisna (osetljiva)

Gramatika bez ograničenja

type-0

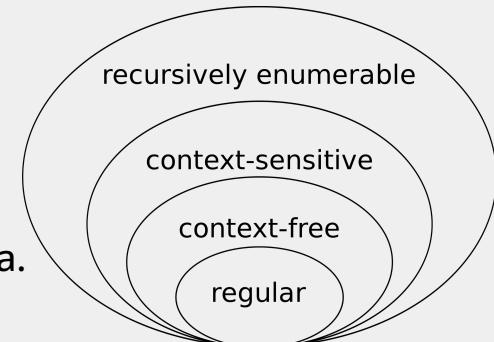
P je skup pravila u obliku

$$\gamma \rightarrow \alpha$$

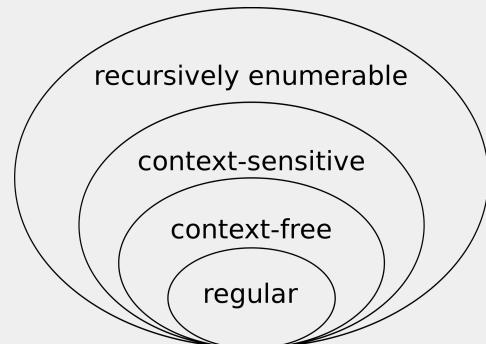
$$\gamma \in (\Sigma \cup N)^+$$

$$\alpha \in (\Sigma \cup N)^*$$

Nema ograničenja za pravila osim da leva strana ne sme biti prazna.



Gramatika	Jezika	Mašina	Pravila
T0	rekurzivno nabrojiv	tjuringova mašina	$\gamma \rightarrow \alpha$
T1: CSG	kontekstno osetljiv	linearno ograničeni automati	$\alpha A \beta \rightarrow \alpha \gamma \beta$
T2: CFG	kontekstno slobodan	potisni automati	$A \rightarrow \alpha$
T3: Reg	regularan	konačni automati	$A \rightarrow a$ ili $A \rightarrow aB$



Svakodnevno programiranje:

- regularni izrazi - koristimo
- regularni izrazi + CFG + CSG (malo) - prevodioci
- tjuringove mašine - programi i računari

Pre nastavka:

Upotreba regularnih i kontekstno slobodnih jezika