

Univerza v Ljubljani
Fakulteta za računalništvo
in informatiko



(Ne)moč končnih avtomatov

15.marec 2014



Moč računskega modela



- Moč računskega modela:
 - katere jezike lahko s tem računskim modelom razpoznamo
 - katere jezike lahko **učinkovito** razpoznamo
- O drugi točki več naslednjič
- V tem sklopu bomo nakazali, da obstajajo jeziki (problemi), ki jih končni avtomati ne znajo razpoznati (rešiti)





Regularni jeziki

- V prejšnjem sklopu smo pokazali, da nedeterministični avtomati prepoznavajo isto množico jezikov kot deterministični
- Tej množici jezikov pravimo regularni jeziki (ker so to natanko isti jeziki kot jih opisujejo regularni izrazi)
- Če želimo pokazati, da ta množica jezikov ne vsebuje vseh možnosti, poiščimo jezik, za katerega ne moremo zapisati KA

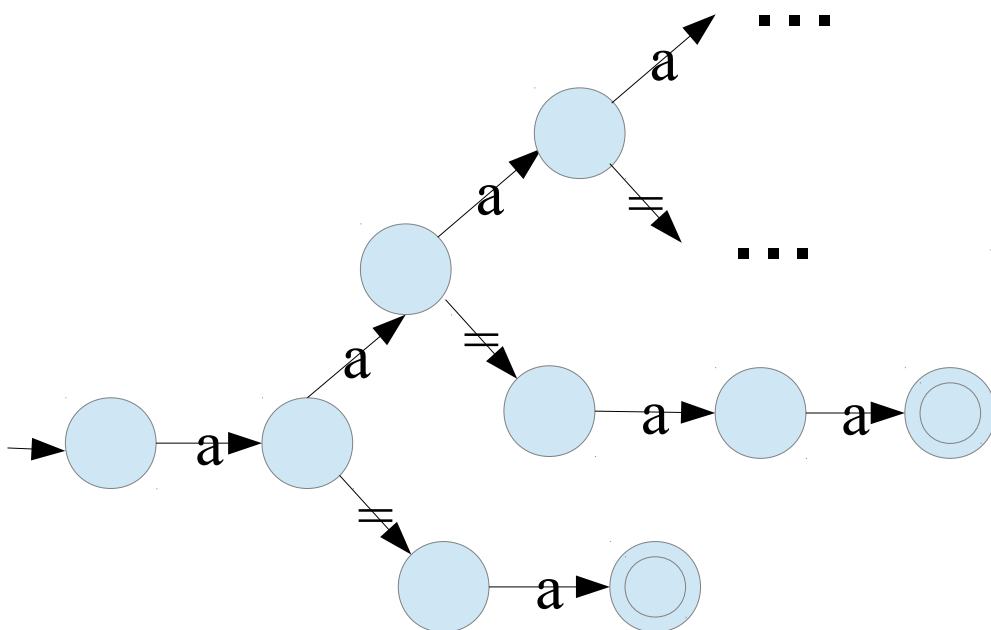




Trd oreh za KA

- Poskusimo sestaviti avtomat, ki bo razpoznal jezik

$$a^n = a^n$$



Ali lahko nekje dodamo zanko?



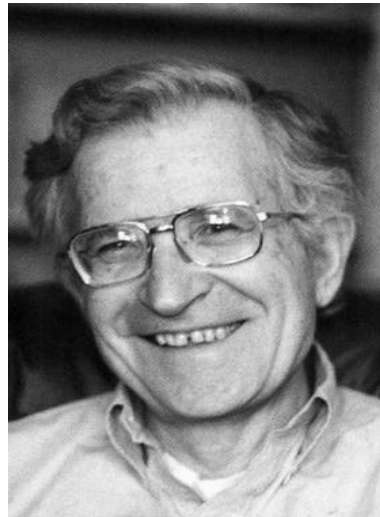
Močnejši formalizem kot so KA

- Končni avtomati predstavljajo formalizem za razpoznavanje jezikov
- V računalništvu pa je pomemben tudi opis in generiranje
- Eden najpomembnejši formalizmov za opis jezikov so **gramatike**.





Zgodovina



Noam Chomsky

(1956)



Gramatike

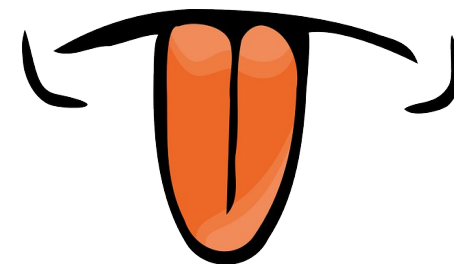
- Chomsky je želel zasnovati formalizem s katerim bi opisali naravne jezike
- Iskal je končen nabor pravil, s katerimi bi lahko zgenerirali vse slovnično pravilne stavke.
- Definiral je veliko različnih vrst gramatik.
- Mi si bomo ogledali zgolj eno vrsto gramatik, t.i. **kontekstno neodvisne gramatike**.





Primer gramatike v naravnem jeziku

Janez bere dobro knjigo.



Stavek → Osebek Povedek Predmet

Osebek → Janez | Metka | Mojca | Kekec

Povedek → bere | grize | meče | boža

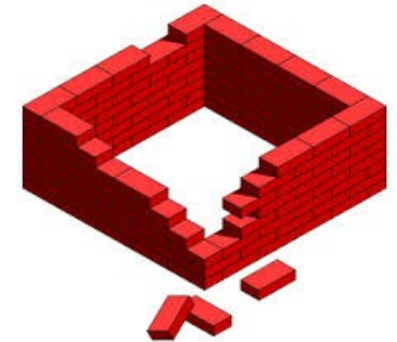
Predmet → Pridevnik Predmet | knjigo | žogo | okno | mizo

Pridevnik → dobro | lepo | trdo | zeleno | nagubano





Gradniki gramatike



- Gramatike vsebujejo 2 vrsti simbolov
 - Spremenljivke (pišemo z velikimi črkami)
 - Končne simbole (pišemo z malimi črkami)
- Pravilom, ki podajajo dovoljene transformacije spremenljivk, pa pravimo produkcije.
- Eni izmed spremenljivk dodelimo status začetnega simbola (iz nje generiramo naš jezik)





Formalna definicija

$$(V, T, P, S)$$

V Končna množica spremenljivk – npr. A, B C, S

T Končna množica končnih simbolov (abeceda) – npr. a,b, 0, 1

$$P = \{ A \rightarrow \alpha : A \in V \wedge \alpha \in (V \cup T)^* \}$$

Končna množica prepisovalnih pravil (produkcij), npr.

$$S \rightarrow abSaS$$

Na levi strani puščice je spremenljivka, na desni pa poljuben niz, sestavljen iz spremenljivk in končnih simbolov.

S Spremenljivka s posebnim statusom, ki ji pravimo začetni simbol





Primer

Gramatika za jezik: $a^n = a^n$

$$V = \{S\} \quad T = \{a, =\} \quad P = \{S \rightarrow aSa, S \rightarrow =\}$$

Ponavadi podajamo samo produkcije, iz konteksta je razvidno vse ostalo.
Npr.

$$S \rightarrow aSa \mid =$$





Zamenjave spremenljivk in izpeljave

Besedo izpeljemo iz začetnega simbola z zamenjavami spremenljivk z desnimi stranmi produkcij. Npr., če

$$A \rightarrow w$$

potem lahko zamenjamo spremenljivko A v nekem nizu uAv (u, v sta poljubna niza) z w kar nam da niz uwv . To zapišemo kot

$$uAv \Rightarrow uwv$$

Niz **$aaa=aaa$** lahko v gramatiki iz prejšnje prosojnice izpeljemo iz začetnega simbola na naslednji način:

$$S \Rightarrow aSa \Rightarrow aaSaa \Rightarrow aaaSaaa \Rightarrow aaa = aaa$$





Jezik gramatike

- Jezik gramatike so vsi nizi (sestavljene iz končnih simbolov), ki jih lahko izpeljemo v poljubnem številu zamenjav.
- Formalno to zapišemo:

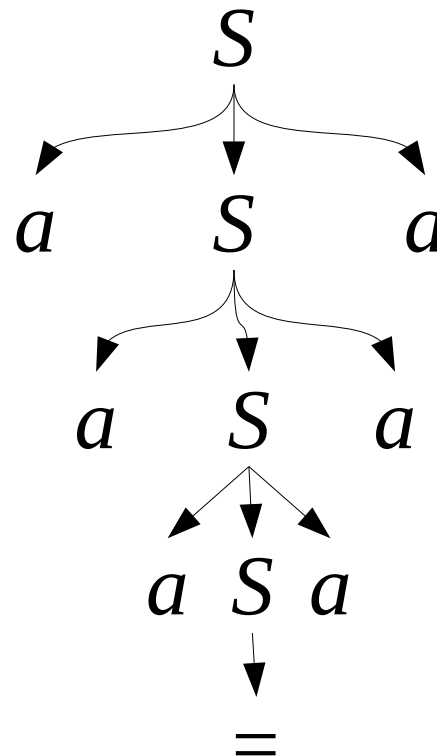
$$L(G) = \{ w \mid S \overset{*}{\Rightarrow} w \}$$





Drevo izpeljave

Včasih zapišemo eno izpeljavo tudi kot drevo, npr. drevo izpeljave za niz $aaa=aaa$:





Dvoumnost gramatik

- Gramatiki pravimo dvoumna, če za nek niz v jeziku obstaja več kot eno drevo izpeljave.
- Dvoumnost igra zelo pomembno vlogo, ker z drevesi izpeljav pogosto vplivamo na »pomen« izpeljanega niza.
- Ogleдали si bomo primer gramatike za aritmetične izraze in kaj mislimo s pojmom »pomen« niza.





Primer – aritmetični izrazi

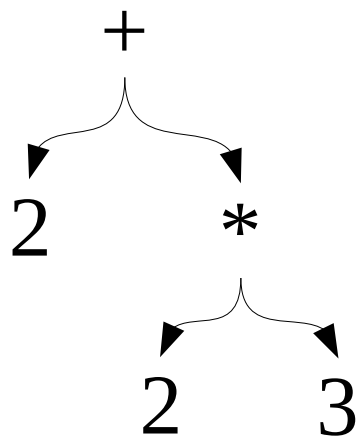
- Zapisati želimo gramatiko, ki bo predstavljala veljavne aritmetične izraze.
- Omejili se bomo na operaciji $+$ in $*$.
- Omejili se bomo tudi na števila 1, 2 in 3.
- Z gramatiko hočemo preveriti pravilnost podanega izraza in tudi definirati kako se bo ta izraz izračunal.



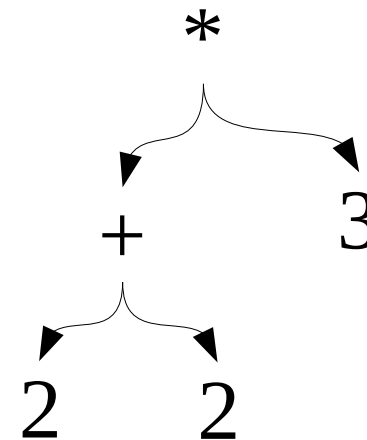


Drevo izraza

$$2+2*3$$



$$2+(2*3)$$



$$(2+2)*3$$

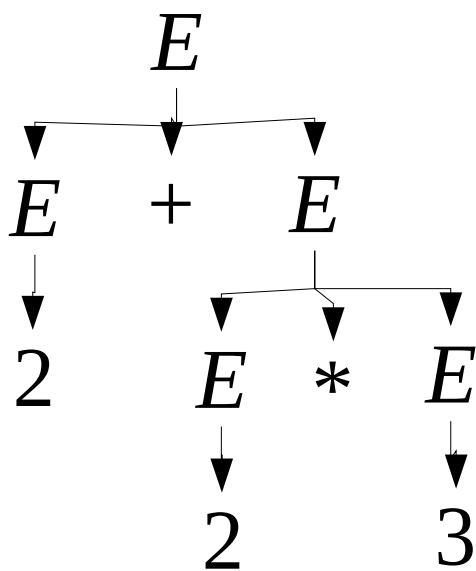




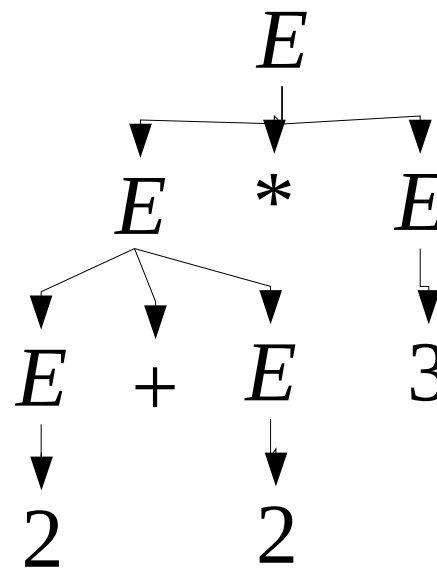
Primer – dvoumna gramatika

$$E \rightarrow E + E \mid E * E \mid 1 \mid 2 \mid 3$$

2+2*3



PRAVILNO



NEPRAVILNO





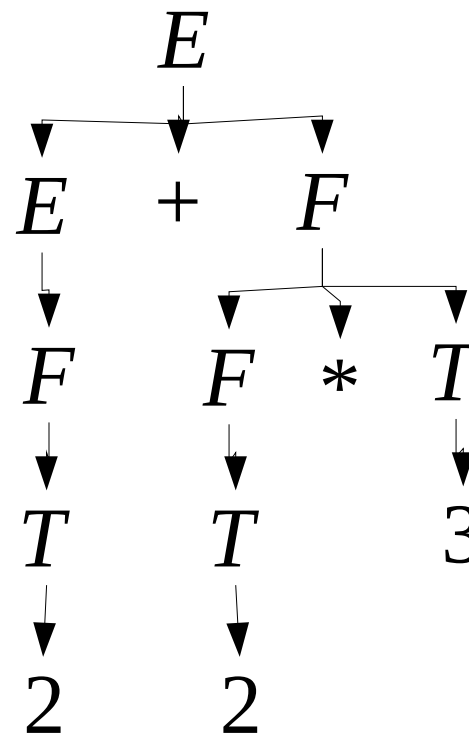
Primer – nedvoumna gramatika

$$E \rightarrow E + F \mid F$$

$$F \rightarrow F * T \mid T$$

$$T \rightarrow (E) \mid 1 \mid 2 \mid 3$$

2+2*3





Uporaba gramatik v računalništvu

- Največ se uporabljajo pri prevajanju programskih jezikov.
- Definicija oblike podatkov (XML).
- Kompresija podatkov.

