Sigurnost u računarskim mrežama

Jedan pristup rasplinutom testiranju softvera
(I deo)

Nemanja Maček

Sadržaj

- Genetski algoritmi
  - Osnovni pojmovi
  - Osnovni genetski algoritam
- Kontekstno nezavisne gramatike
  - Osnovni pojmovi
  - Formalne definicije
  - Primeri

Napomene

- Na ovom predavanju ćemo razmotriti:
  - Osnove genetskih algoritama.
  - Osnove kontekstno nezavisnih gramatika.
- Ova dva koncepta ćemo potom primeniti u okviru teme fazi testiranje softvera.

Osnovni pojmovi

- **Genetski algoritmi** su heuristička metoda optimizacije inspirisana evolutivnim procesom u prirodi.
- Osnovna ideja:
  - Postoji populacija jedinki.
  - Neke jedinke su bolje prilagođene okolini od drugih.
  - Bolje prilagođene jedinke će sa većom verovatnoćom ostaviti potomstvo.
  - Potomstvo nasleđuje kombinaciju roditeljskih gena (a može i da mutira), pa će svaka naredna generacija biti bolje prilagođena okolini od prethodne.

Osnovni pojmovi

Hromozomi.

- Veoma uprošćeni osvrt na biološku terminologiju:
  - Hromozomom je niz gena.
  - Svaki gen se nalazi na određenoj poziciji u hromozomu.
  - Moguće “vrednosti” gena su – aleli.

Osnovni pojmovi

Hromozomi.

- U kontekstu genetskih algoritama, pod hromozomom podrazumevamo potencijalno rešenje nekog problema, i često ga predstavljamo nizom bitova.
- U ovakvoj reprezentaciji:
  - Geni su predstavljeni bitovima (ili kratkim blokovima bitova).
  - Aleli su predstavljeni skupom {0,1}.
- Pod **ukrštanjem hromozoma** podrazumevamo razmenu gena (npr., blokova bitova) između dva hromozoma.
- Pod **mutacijom hromozoma** podrazumevamo promenu vrednosti slučajno izabranih gena.

Osnovni pojmovi

Elementi genetskih algoritama.

- Iako ne postoji stroga definicija genetskog algoritma, svi “genetski algoritmi” poseduju neke zajedničke elemente:
  - Populaciju hromozoma.
  - Selekciju jedinki u skladu sa stepenom njihove prilagođenosti okolini (engl. *fitness*).
    - Ovde prilagođenost podrazumeva meru kvaliteta jedinke.
  - Ukrštanje jedinki (engl. *crossover*) i produkciju potomstva.
  - Slučajne mutacije potomstva (engl. *mutation*).

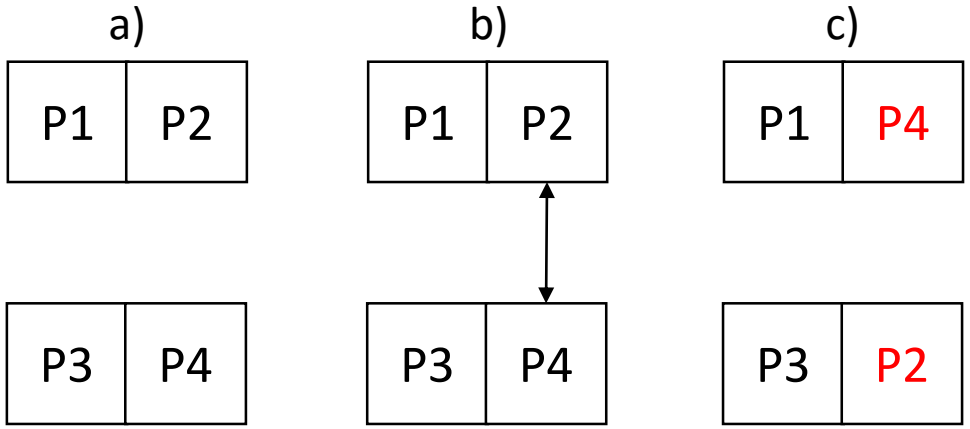
Osnovni genetski algoritam

Osnovni algoritam.

1. Generiše se populacija od *n* hromozoma dužine *l* bitova (svaki hromozom predstavlja potencijalno rešenje problema koji se razmatra).
2. Za svaki hromozom *x* u populaciji se izračuna stepen prilogođenosti $f(x)$.
3. Koraci 3.a-3.c se ponavljaju dok se ne kreira *n* potomaka:
   a) Iz populacije se izaberu dva roditelja, pri čemu se bolje prilagođene jedinke biraju sa većom verovatnoćom.
   b) Sa verovatnoćom p_c se vrši ukrštanje izabranih hromozoma na slučajno izabranoj poziciji gena – i generišu se dva potomka.
   c) Sa verovatnoćom p_m se vrši mutiranje potomaka, a potom se smeštaju u novu populaciju.
4. Početna populacija hromozoma se zameni novom.
5. Skok na korak 2.

Osnovni genetski algoritam

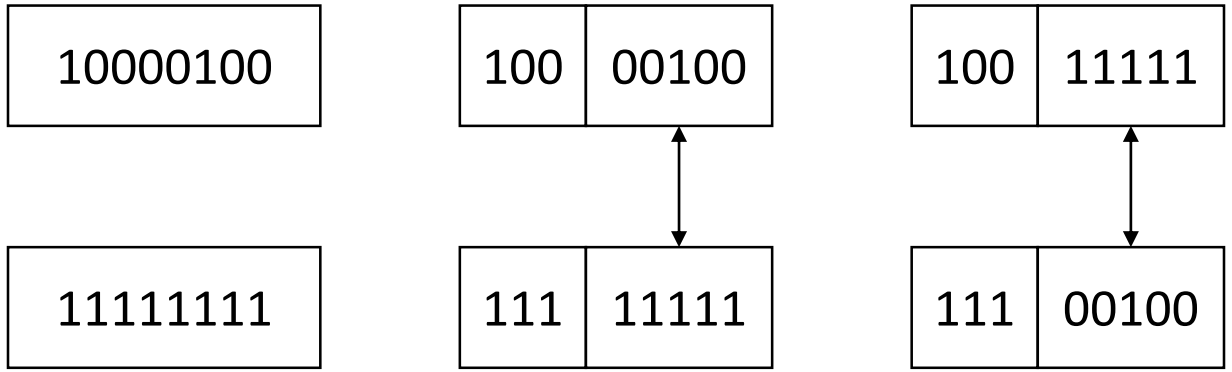
Ukrštanje.



a) Za izabrane roditelje se na slučajni način izabere pozicija gena koji ih deli na dva dela.
b) Ukrštanje se vrši tako što roditelji razmene delove koji se nalaze posle izabranog gena.
c) Na taj način se kreiraju dva potomka.

Osnovni genetski algoritam

Ukrštanje – primer.



- Za roditelje 10000100 i 11111111 koji razmenjuju delove hromozoma posle trećeg gena se dobijaju dva potomka:
  - 10011111 i
  - 11100100.

Osnovni genetski algoritam

Mutacija.

- Prilikom mutacije hromozoma se komplementira vrednost slučajno izabranog gena (tj., 1 prelazi u 0, a 0 u 1).
- Npr, mutacijom hromozoma 00000100 na drugoj poziciji se dobija hromozom 01000100.

Osnovni genetski algoritam

Početna populacija.

- Neka početna populacija sadrži 4 hromozoma dužine 8 bita.
- Neka funkcija prilagođenosti hromozoma vraća vrednost jednaku broju jedinica u hromozomu.

Hromozom	Niz	Prilagođenost
A	00000110	2
B	11101110	6
C	00100000	1
D	00110100	3

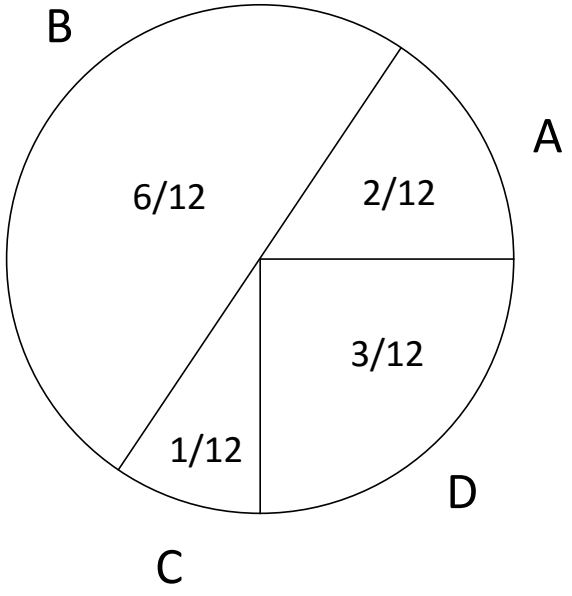
- Neka su:
  - Verovatnoća ukrštanja $p_c = 0,7$
  - Verovatnoća mutacije $p_m = 0,001$.

Osnovni genetski algoritam

Rulet-selekcija.

- Bolje prilagođeni hromozomi se biraju sa većom verovatnoćom.
- Jedan od mogućih načina za selekciju hromozoma je **rulet-selekcija**, kod koje je verovatnoća biranja hromozoma proporcionalna njegovoj prilagođenosti.

Hromozom	Prilagođenost
A	2
B	6
C	1
D	3
	$\Sigma = 12$



Osnovni genetski algoritam

Ukrštanje hromozoma.

- Pošto u populaciji imamo 4 hromozoma, rulet-selekcija će biti primenjena 4 puta – biraju se dva para roditelja, pri čemu svaki par generiše po dva potomka.
- Pretpostavimo da su prva dva selektovana hromozoma B i D, i da se ukrštaju posle prvog gena.
  - Dobijaju se dva potomka: E (10110100) i F (01101110).
- Pretpostavimo da su druga dva selektovana hromozoma B i C, i da se oni ne ukrštaju – tj., njihovi potomci su isti kao i roditelji.

Osnovni genetski algoritam

Ukrštanje hromozoma.

- Pošto u populaciji imamo 4 hromozoma, rulet-selekcija će biti primenjena 4 puta – biraju se dva para roditelja, pri čemu svaki par generiše po dva potomka.
- Pretpostavimo da su prva dva selektovana hromozoma B i D, i da se ukrštaju posle prvog gena.
  - Dobijaju se dva potomka: E (10110100) i F (01101110).
- Pretpostavimo da su druga dva selektovana hromozoma B i C, i da se oni ne ukrštaju – tj., njihovi potomci su isti kao i roditelji.
- **Pažnja**:
  - Selektovani hromozomi se ukrštaju sa verovatnoćom p_c, što znači da se ponekad neće ukrštati.
    - Vrednost p_c =0,7 (koju koristimo u ovom primeru) znači da se, u proseku, 30% izabranih parova hromozoma neće ukrštati.
  - Dozvoljeno je više puta birati isti hromozom (u ovom primeru – B).

Osnovni genetski algoritam

Mutacija potomaka.

- Potomci mutiraju sa verovatnoćom p_m.
- U ovom primeru je $p_m = 0{,}001$, što znači da će, u proseku, svaki hiljaditi hromozom mutirati.
- Pretpostavimo da:
  - Potomak E (10110100) mutira na 6. genu – tj. dobija se mutirani hromozom E^1 (10110000).
  - Potomci C i F ne mutiraju.
  - Potomak B (11101110) mutira na 1. genu – tj. dobija se mutirani hromozom B^1 (01101110).

Osnovni genetski algoritam

Početna populacija		
Hromozom	Niz	Prilagođenost
A	00000110	2
B	11101110	6
C	00100000	1
D	00110100	3

Nova populacija		
Hromozom	Niz	Prilagođenost
E^1	10110000	3
F	01101110	5
C	00100000	1
B^1	01101110	5

Osnovni genetski algoritam

Procena napretka.

Početna populacija		Nova populacija	
Hromozom	Prilagođenost	Hromozom	Prilagođenost
A	2	E^1	3
B	6	F	5
C	1	C	1
D	3	B^1	5

- Iako nova populacija ne sadrži najbolji hromozom iz početne populacije (prilagođenost hromozoma B je bila 6), prosečna prilagođenost hromozoma u novoj populaciji (14/4=3,5) je bolja nego u početnoj populaciji (12/4=3).

Osnovni genetski algoritam

Napredak kroz generacije.

- Ako bismo nastavili da sukcesivno kreiramo nove populacije, prosečna prilagođenost hromozoma bi rasla.
  - $\text{populacija}_1 \rightarrow \text{populacija}_2 \rightarrow \ldots \text{populacija}_k$
- U jednoj od populacija (k) bi se pojavio "optimalni" hromozom koji sadrži sve jedinice.

Završne napomene o genetskim algoritmima

- U ovom izlaganju predavanju je predstavljen samo osnovni genetski algoritam: hromozomi su predstavljeni nizom bitova, a skup alela je ograničen na {0,1}.
- Postoje mnogobrojne varijacije ovog algoritma koje odstupaju od osnove verzije po načinu predstavljanja, selektovanja, ukrštanja i mutiranja hromozoma.

Kontekstno nezavisne gramatike

- U ovom delu izlaganja ćemo razmotriti osnove kontekstno nezavisnih gramatika (engl. *context free grammars*).
  - U literaturi na srpskom jeziku ćete naći i frazu kontekstno slobodne gramatike.
- Krenimo od primera …

Osnovni pojmovi

- Primer (uprošćene) kontekstno slobodne gramatike:
  - $S \rightarrow NP\ VP$
  - $VP \rightarrow V\ NP$
  - $NP \rightarrow$ pas | mačku | dečko | jabuku
  - $V \rightarrow$ juri | jede

Neterminalni i terminalni simboli

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

- *S*, *NP*, *VP*, *V* su **neterminalni simboli** koji predstavljaju "sintaksne" entitete u posmatranom jeziku.
  - Ovi simboli su "neterminalni" jer se mogu dalje izvoditi u druge entitete.
- pas, mačku, dečko, jabuku, juri, jede su terminalni simboli (tj. konačni, koji ne generišu druge).
- Neterminalni simboli se uobičajeno pišu velikim slovima, a terminalni malim.

Produkciona pravila

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

- $NP \rightarrow$ pas je produkciono pravilo koje kaže da se iz *NP* može izvesti imenica “pas”.
- $NP \rightarrow$ pas | mačku | dečko | jabuku je produkciono pravilo koje kaže da se iz *NP* mogu izvesti imenice “pas”, “mačku”, “dečko”, “jabuku”.

Produkciona pravila

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

- $S \rightarrow NP\ VP$ je produkciono pravilo koje kaže da se *S* sastoji iz dva dela, *NP* i *VP*, pri čemu *NP* dolazi pre *VP*.
- $VP \rightarrow V\ NP$ je produkciono pravilo koje kaže da se *VP* sastoji iz dva dela, *V* i *NP*, pri čemu *V* dolazi pre *NP*.

Izvođenje rečenice

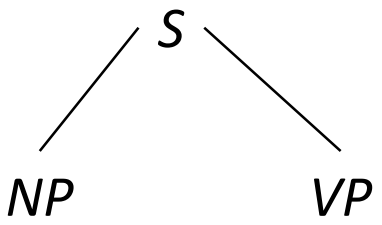
- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

- Izvođenje rečenice uvek počinjemo od simbola *S*.

Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

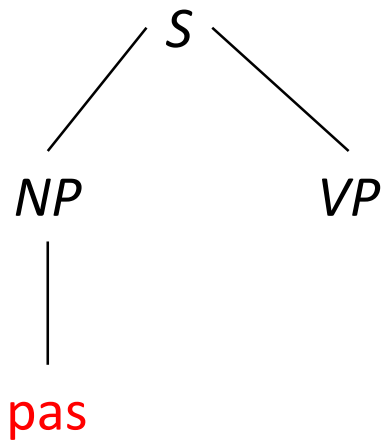
- Pravilo 1.



Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

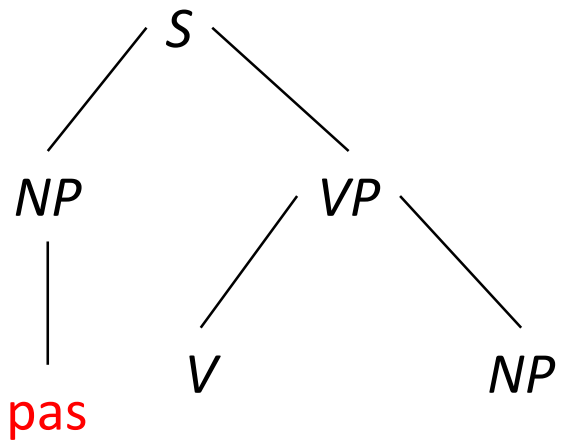
- Pravilo 3.



Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

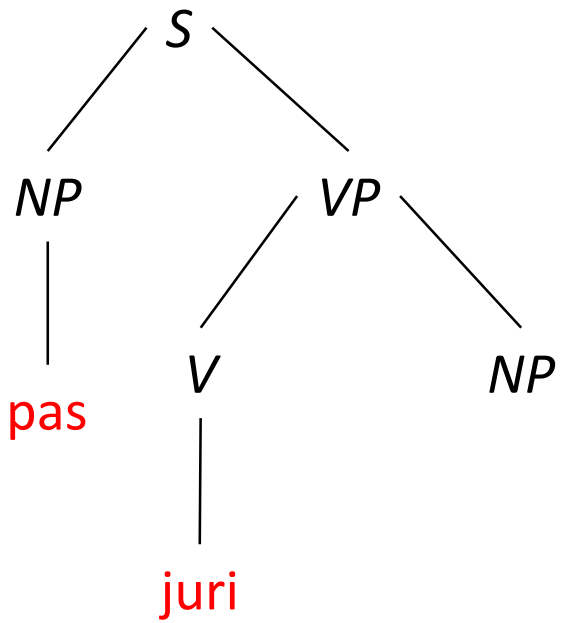
- Pravilo 2.



Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

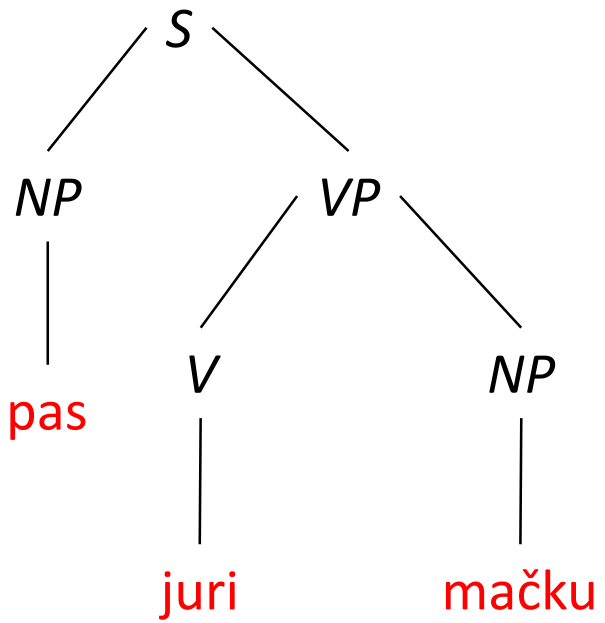
- Pravilo 4.



Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

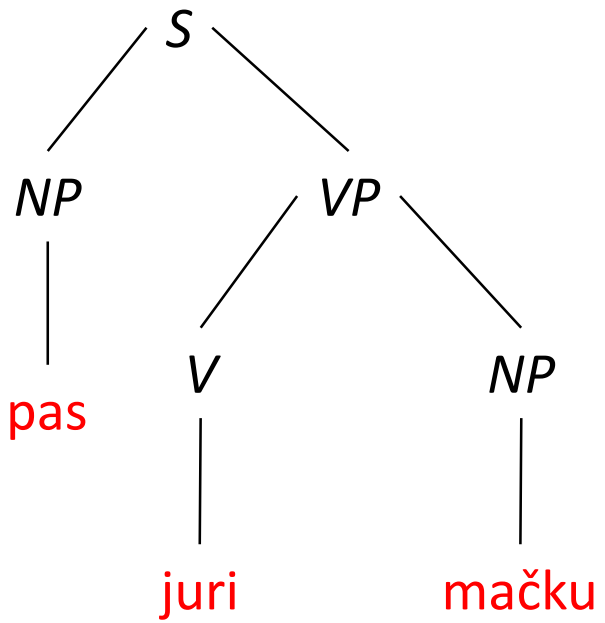
- Pravilo 3 (ponovo).



Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

- Rečenica “pas juri mačku” pripada jeziku definisanom datom gramatikom, jer postoji **sintaksno stablo** koje je izvodi.



Izvođenje rečenice

- $S \rightarrow NP\ VP$
- $VP \rightarrow V\ NP$
- $NP \rightarrow$ pas | mačku | dečko | jabuku
- $V \rightarrow$ juri | jede

- “dečko jede jabuku”, “dečko jede mačku”, “jabuku jede dečko” i “mačku jede jabuku” su primeri rečenica koje pripadaju datom jeziku (iako poslednja ne odgovara gramatici srpskog jezika).
- “mačka juri psa” i “dečko jabuku jede” su primeri rečenica koje ne pripadaju datom jeziku (iako odgovaraju gramatici srpskog jezika) – jer ne postoji sintaksno stablo za njihovo izvođenje.

Formalne definicije

- **Definicija kontekstno nezavisne gramatike**.
- Kontekstno nezavisna gramatika je uređena četvorka
  $G = \{N, \Sigma, P, S\}$ gde su:
  - N – skup neterminalnih simbola
  - Σ – skup terminalnih simbola (važi: $N \cap \Sigma = \emptyset$)
  - P – skup produkcionih pravila oblika: $A \rightarrow \alpha$, gde su:
    - A – neterminalni symbol
    - α – string iz beskonačnog skupa stringova $(N \cup \Sigma)^*$,
  - S – početni simbol.

Formalne definicije

- **Direktno izvođenje**.
- Neka su:
  - $A \rightarrow \beta$ pravilo iz skupa P
  - α i γ stringovi iz skupa $(N \cup \Sigma)^*$, pri čemu ovi stringovi mogu da budu i prazni
- Onda kažemo da $\alpha A \gamma$ direktno izvodi $\alpha \beta \gamma$:
  - $\alpha A \gamma \Rightarrow \alpha \beta \gamma$

Formalne definicije

- **Izvođenje**.
- Neka su α_1, α_2, … α_m, m>1 stringovi iz skupa $(N \cup \Sigma)^*$ takvi da važi:
  - $\alpha_1 => \alpha_2$, $\alpha_2 => \alpha_3$, …, $\alpha_{m-1} => \alpha_m$.
- Onda kažemo da α_1 izvodi α_m:
  - $\alpha_1 =*> \alpha_m$
- Gramatika G generiše jezik L:
  - $L = \{w \mid w \in \Sigma^*, S =*> w\}$, tj. skup stringova terminalnih simbola koji se mogu izvesti iz S.

Primeri

- **Primer 1**.
- Jezik $L = \{w \in \{a, b\}^* \mid w$ je izraz tipa $aa\ldots abb\ldots.b$ pri čemu se a i b ponavljaju n puta, $n \geq 1\}$.
- Alternativni zapis: $L = \{a^n b^n \mid n \geq 1\}$.

Primeri

- **Primer 1**.
- Jezik $L = \{w \in \{a, b\}^* \mid w$ je izraz tipa $aa\ldots abb\ldots.b$ pri čemu se a i b ponavljaju n puta, $n \geq 1\}$.
- Alternativni zapis: $L = \{a^n b^n \mid n \geq 1\}$.
- Gramatika koja definiše ovaj jezik:
  - $S \rightarrow aSb$
  - $S \rightarrow ab$

Primeri

- **Primer 2**.
- Jezik $L = \{a^n b^n \mid n \geq 0\}$.

Primeri

- **Primer 2**.
- Jezik $L = \{a^n b^n \mid n \geq 0\}$.
- Gramatika koja definiše ovaj jezik:
  - $S \rightarrow aSb$
  - ~~$S \rightarrow ab$~~ ovo pravilo više nije potrebno
  - $S \rightarrow \varepsilon$, pri čemu je ε prazan string
- Alternativni zapis gramatike:
  - $S \rightarrow aSb \mid \varepsilon$

Primeri

- **Primer 3**.
- Jezik $L = \{w \in \{a, b\}^* \mid w \text{ je palindrom}\}$.
  (Palindrom je string koji se isto piše sleva nadesno i sdesna nalevo).

Primeri

- **Primer 3**.
- Jezik $L = \{w \in \{a, b\}^* \mid w \text{ je palindrom}\}$.
  (Palindrom je string koji se isto piše sleva nadesno i sdesna nalevo).
- Gramatika koja definiše ovaj jezik:
  - $S \rightarrow aSa \mid bSb \mid a \mid b \mid \varepsilon$

Primeri

- **Primer 4**.
- Koji jezik generiše sledeća gramatika?
  - $S \rightarrow aSa \mid B$
  - $B \rightarrow bB \mid \varepsilon$

Primeri

- **Primer 4**.
- Koji jezik generiše sledeća gramatika?
  - $S \rightarrow aSa \mid B$
  - $B \rightarrow bB \mid \varepsilon$
- $L = \{w \in \{a, b\}^* \mid w$ je izraz tipa $aa…abb….baa…a$ pri čemu se a i b ponavljaju n i m puta, respektivno, $n,m \geq 0\}$.
- Neformalni opis: svaki string ovog jezika počinje i završava se podjednako dugačkim sekvencama simbola a (koje mogu biti i prazne), a između njih se nalazi sekvenca simbola b proizvoljne dužine (koja takođe može biti prazna).

Primeri

- **Primer 5**.
- Koji jezik generiše sledeća gramatika?
  - $S \rightarrow a \mid b$
  - $S \rightarrow (S)$
  - $S \rightarrow S + S$
  - $S \rightarrow S - S$
  - $S \rightarrow S \times S$
  - $S \rightarrow S \div S$

Primeri

- **Primer 5**.
- Koji jezik generiše sledeća gramatika?
  - $S \rightarrow a \mid b$
  - $S \rightarrow (S)$
  - $S \rightarrow S + S$
  - $S \rightarrow S - S$
  - $S \rightarrow S \times S$
  - $S \rightarrow S \div S$
- Jezik sadrži stringove koji odgovaraju aritmetičkim izrazima sa operacijama +, –, ×, ÷ nad argumentima *a*, *b*.

Završne napomene o kontekstno nezavisnim gramatikama

- Šta znači kontekstno nezavisna?
  - Pojam konteksta u kontekstno nezavisnim gramatikama nema veze sa uobičajenim značenjem konteksta u prirodnom jeziku.
  - Gramatika je “kontekstno nezavisna” ukoliko se sa leve strane produkcionih pravila nalazi tačno jedan neterminalni simbol (i ništa osim njega).

Literatura

1. M. Gnjatović, D. Stefanović (2018): Izabrane teme iz bezbednosti i sigurnosti informacionih sistema. Fakultet tehničkih nauka, Univerzitet u Novom Sadu.
2. Embleton, S., Sparks, S., Cunningham, R. (2006) Sidewinder: An Evolutionary Guidance System For Malicious. Input Crafting, Black Hat USA 2006, http://www.blackhat.com/presentations/bh-usa-06/BH-US-06-Embleton.pdf.

Hvala na pažnji

Pitanja su dobrodošla.