Kontekstno slobodni jezici

školska 2018/2019

Uvod

- Kontektstno slobodni jezici su šira klasa jezika u odnosu na regularne jezike
- Kontekstno slobodne gramatike igraju glavnu ulogu u razvoju kompajlera još od 1960tih godina
- Danas se u velikoj meri koriste za opise formata dokumenata (*document-type definition*, DTD) koji se koriste u XML zajednici za razmenu informacija na vebu

Definicija

Дефиниција

Gramatika $\mathbb{G} = (\Gamma, \Sigma, S, \mathcal{P})$ je kontekstno-slobodna ako su sva njena pravila oblika $X \to w$, za neko $X \in \Gamma$ i neku reč $w \in (\Gamma \cup \Sigma)^*$.

1. Za pravilo $X \to w$, kažemo da je X glava ovog pravila, a w telo pravila.
2. Ukoliko u gramatici postoji više pravila sa istom glavom:

$$X \to w_1,\ X \to w_2, \ldots, X \to w_k, \text{ kraće zapisujemo}$$

$$X \to w_1 \mid w_2 \mid \cdots \mid w_k.$$

3. Pravilo $X \to w$ je linearno ako je $w = uYv$, za neke $u, v \in \Sigma^*$ i $Y \in \Gamma$.
4. Pavila oblika $X \to Y$, $X, Y \in \Gamma$, nazivaju se lančasta pravila.
5. Pravila oblika $X \to \varepsilon$, $X \in \Gamma$ nazivamo pravilima brisanja.

Definicija

Дефиниција

Jezik $L \subseteq \Sigma^*$ *je kontekstno-slobodan ako postoji kontekstno slobodna gramatika* $\mathbb{G} = (\Gamma, \Sigma, S, \mathcal{P})$ *takva da je* $L = L(\mathbb{G})$.

- Svaki regularan jezik je kontekstno slobodan, a obrnuto nije tačno.

Drvo izvodjenja

- Prikazivanje izvodjenja drvetom potpuno je prirodno - dinamiku izvodjenja odredjuju promenljive kojih može biti više od jedne sa desne strane pravila, a nije važan redosled kojim će biti aktivirane.
- Ako je $\mathbb{G} = (\Gamma, \Sigma, S, \mathcal{P})$ kontekstno slobodna gramatika i $X \in \Gamma$, X-drvo izvodjenja je drvo čiji su čvorovi označeni elementima iz $\Gamma \cup \Sigma \cup \{\varepsilon\}$ i važi:
  - koren drveta je označen sa X,
  - listovi su označeni terminalima ili ε, a unutrašnji čvorovi promenljivama,
  - ako je neki unutrašnji čvor označen promenljivom Y, a njegovi neposredni sledbenici simbolima $s_1, \dots, s_k \in \Gamma \cup \Sigma$, onda $Y \to s_1 \cdots s_k \in \mathcal{P}$,
  - ako je list označen sa ε neposredni sledbenik nekog unutrašnjeg čvora Y, onda je on i jedini neposredni sledbenik tog čvora.
- Rezultat nekog drveta izvodjenja jeste reč koja se dobija dopisivanjem sleva nadesno listova tog drveta.
- Svaki unutrašnji čvor nekog drveta izvodjenja zajedno sa svim svojim sledbenicima obrazuje novo drvo izvodjenja.

Drvo izvodjenja

- Svakom izvodjenju u kontekstno slobodnoj gramatici odgovara jedinstveno drvo izvodjenja, dok obrnuto nije tačno.
- Ako se ograničimo samo na tzv. leva izvodjenja, dobijamo obostrano jednoznačnu korespondenciju izmedju izvodjenja i drveta.
- Leva izvodjenja su ona u čijem je svakom koraku $w \to_{\mathbb{G}} w'$ primenjeno pravilo čija je glava najlevlja promenljiva iz w.
- Može se dokazati da svaku kontekstnu slobodnu gramatiku $\mathbb{G} = (\Gamma, \Sigma, S, \mathcal{P})$ važi:

  $S \to_{\mathbb{G}}^{*} w$ akko postoji levo izvodjenje reči w iz S.
- Do istih zaključaka bismo došli da smo posmatrali desna izvodjenja.

Dvosmislenost jezika

Дефиниција

Kontekstno slobodna gramatika $\mathbb{G}$ je dvosmislena ako postoji reč $w \in L(\mathbb{G})$ koja ima dva različita drveta izvodjenja u $\mathbb{G}$. U suprotnom, gramatika je nedvosmislena.

Дефиниција

Kontekstno slobodni jezik je suštinski dvosmislen ako je dvosmislena svaka gramatika koja ga generiše.

Upotreba kontekstno slobodnih gramatika

- Kontekstno slobodne gramatike igraju glavnu ulogu u razvoju kompajlera od 1960tih
- Parseri su programi kojima se proverava ispravna struktuiranost ulaznih podataka i omogućava njihova dalja obrada - YACC
- Većina markap jezika (poput HTML, XML) je kontektsno slobodna, pa se definisanjem gramatičkih pravila parserom lako proverava i prihvata sadržaj opisan ovim jezicima

Redukovana gramatika

- Svaka kontekstno slobodna gramatika može se transformisati u gramatiku koja generiše isti jezik ali čija su pravila veoma jednostavnog oblika.
- Neka je $\mathbb{G} = (\Gamma, \Sigma, S, \mathcal{P})$ kontekstno slobodna gramatika. Promenljiva $X \in \Gamma$ je
  - završavajuća ako postoji reč $w \in \Sigma^*$ takva da $X \to_{\mathbb{G}}^* w$;
  - dostižna ako postoje $u, v \in (\Gamma \cup \Sigma)^*$ takve da $S \to_{\mathbb{G}}^* uXv$.

Дефиниција

Kontekstno slobodna gramatika je redukovana ako su sve njene promenljive završavajuće i dostižne.

Redukovana gramatika

- Za kontekstno slobodnu gramatiku $\mathbb{G} = (\Gamma, \Sigma, S, \mathcal{P})$ mogu se efektivno odrediti:
    - skup $\mathcal{Z}$ svih završavajućih promenljivih i
    - skup $\mathcal{D}$ svih dostižnih promenljivih.
- Definišimo skupove $\mathcal{Z}_i$ i $\mathcal{D}_i$, $i \geqslant 0$, induktivno, na sledeći način:
  $\mathcal{Z}_0 = \{X \in \Gamma \mid (\exists w \in \Sigma^*) X \to w \in \mathcal{P}\}$,
  $\mathcal{Z}_{i+1} = \{X \in \Gamma \mid (\exists w \in (\mathcal{Z}_i \cup \Sigma)^*) X \to w \in \mathcal{P}\}$, $i \geqslant 0$;
  $\mathcal{D}_0 = \{S\}$,
  $\mathcal{D}_{i+1} = \mathcal{D}_i \cup \{X \in \Gamma \mid (\exists Y \in \mathcal{D}_i)(\exists u, v \in (\Gamma \cup \Sigma)^*) Y \to uXv \in \mathcal{P}\}$, $i \geqslant 0$.
- Tada važe sledeće osobine:

$\mathcal{Z}_0 \subseteq \mathcal{Z}_1 \subseteq \mathcal{Z}_2 \subseteq \cdots$	$\mathcal{D}_0 \subseteq \mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \cdots$
$\mathcal{Z}_i = \mathcal{Z}_{i+1} \Rightarrow \mathcal{Z}_i = \mathcal{Z}_{i+1} = \mathcal{Z}_{i+2} = \cdots$	$\mathcal{D}_i = \mathcal{D}_{i+1} \Rightarrow \mathcal{D}_i = \mathcal{D}_{i+1} = \mathcal{D}_{i+2} = \cdots$
$\mathcal{Z}_i = \mathcal{Z}_{i+1}$, za $i = \lvert\Gamma\rvert - 1$	$\mathcal{D}_i = \mathcal{D}_{i+1}$, za $i = \lvert\Gamma\rvert - 1$
$\mathcal{Z} = \bigcup_{i \geqslant 0} \mathcal{Z}_i$	$\mathcal{D} = \bigcup_{i \geqslant 0} \mathcal{D}_i$

  odakle sledi da se $\mathcal{Z}$ i $\mathcal{D}$ mogu efektivno odrediti.

Redukovana gramatika

Napomena. Problem $L(\mathbb{G}) \neq \emptyset$ je odlučiv, jer je ekvivalentan sa $S \in \mathcal{Z}$.

Теорема

Za svaku kontekstno slobodnu gramatiku $\mathbb{G}$, takvu da je $L(\mathbb{G}) \neq \emptyset$, postoji redukovana kontekstno slobodna gramatika $\mathbb{G}^{\mathrm{r}}$ takva da je $L(\mathbb{G}) = L(\mathbb{G}^{\mathrm{r}})$.

Normalna forma Čomskog

Дефиниција

Kontekstno slobodna gramatika $(\Gamma, \Sigma, S, \mathcal{P})$ *je u normalnoj formi Čomskog ako su sva njena pravila oblika:*

- $X \to YZ$, $X, Y, Z \in \Gamma$,
- $X \to x$, $X \in \Gamma$, $x \in \Sigma$, *ili*
- $S \to \varepsilon$,

i pri tome, ako se $S \to \varepsilon$ *pojavljuje medju pravilima, onda se* S *ne pojavljuje sa desne strane nijednog pravila.*

Теорема

Za svaku kontekstno slobodnu gramatiku $\mathbb{G}$ *postoji kontekstno slobodna gramatika* $\mathbb{G}^{\mathrm{CH}}$ *u normalnoj formi Čomskog takva da je* $L(\mathbb{G}) = L(\mathbb{G}^{\mathrm{CH}})$.

Normalna forma Čomskog

Lema

Ako je G kontekstno slobodna gramatika u normalnoj formi Čomskog, onda za svaku reč $w \in L(G)$ takvu da je $|w| \geqslant 1$ potrebno je tačno $2|w| - 1$ koraka za bilo koje izvodjenje reči w u gramatici G.