

Bezkontextové gramatiky a zásobníkové automaty, formální popis, ekvivalence

Thursday, May 30, 2013 8:39 AM

<https://class.coursera.org/compilers/lecture/38>

Bezkontextové gramatiky

Bezkontextové gramatiky (BKG) jsou gramatiky typu 2 podle Chomského hierarchie.

Gramatika je bezkontextová, **tvoří-li levou stranu všech přepisovacích pravidel právě jeden neterminální symbol**

Skládají se tedy z pravidel

$$A \rightarrow \gamma$$

kde A je právě jeden neterminál a

γ

je řetězec terminálů a neterminálů. Pravidlo

$$S \rightarrow \epsilon$$

je povoleno, pokud se S nevyskytuje na pravé straně žádného pravidla. Jazyky generované touto gramatikou jsou **rozpoznatelné nedeterministickým zásobníkovým automatem**.

Skládají se z **terminalů, neterminálů, počátečního symbolu a přepisovacích (produkčních) pravidel**; **G = (N,T,P,S)**.

- T = Terminály – jde o názvy tokenů (klíčová slova *if, else*, symboly „(“, „)“ atd.)
- N = Neterminály – syntaktické proměnné, pomáhají definovat jazyk generovaný gramatikou; zavádějí hierarchickou strukturu jazyka, která je klíčová pro syntaktickou analýzu
- S = Počáteční symbol – jeden z neterminálů
- P = přepisovací pravidla - ve tvaru

$$A \rightarrow \gamma, \text{ kde } A \in N \text{ a } \gamma \in N \cup T$$

Příklad

Jazyk

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

pro

$$n \geq 0$$

, takovýto jazyk není rozpoznatelný konečným automatem, zásobníkovým ano („konečný automat neumí počítat“). U programovacích jazyků by to třeba znamenalo, že není možné závorkovat (vnořovat kód) do libovolné úrovně.

Pro tento jazyk by platilo:

$$N = \{S\}$$

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

$$P = \{S \rightarrow 0S1, S \rightarrow \epsilon\}$$

$$S = \{S\}$$

Zásobníkový automat

Formálně je zásobníkový automat definován jako **uspořádaná sedmice (Q,T,G, δ , q_0 , z_0 , F)**, kde:

- Q je konečná množina vnitřních stavů,
- T je konečná vstupní abeceda,
- G je konečná abeceda zásobníku,
- δ je tzv. přechodová funkce, popisující pravidla činnosti automatu (jeho program), je definováno jako zobrazení

$$Q \times (T \cup \{e\}) \times G^* \text{ do } Q \times G^*$$

- q_0 je počáteční stav,
- z_0 popisuje symboly uložené na počátku v zásobníku,
- F je množina přijímajících stavů,

$$F \subseteq Q$$

Je vidět, že zásobníkový automat se v podstatě skládá z konečného automatu, který má navíc k dispozici potenciálně nekonečné množství paměti ve formě zásobníku. Obsah tohoto zásobníku ovlivňuje činnost automatu tím, že vstupuje jako jeden z parametrů do přechodové funkce.

Zásobníkový automat se od konečného automatu liší ve dvou směrech:

1. Využívá vršek zásobníku při rozhodování jaký přechod provést.
2. Může manipulovat se zásobníkem jako součást provádění přechodu.

Popis činnosti automatu

Na počátku se automat nachází v definovaném počátečním stavu a zásobník obsahuje pouze počáteční symboly. Dále v každém kroku podle aktuálního stavu, symbolů na vrcholu zásobníku a symbolu na vstupu provede přechod, při kterém může vyjmout ze zásobníku několik symbolů, vložit místo nich jiné a na vstupu přečíst další symbol. Toto se opakuje.

Po dokončení činnosti (po přečtení celého vstupu, pokud do té doby nedojde k chybě) je rozhodnuto, jestli automat vstupní řetězec přijal. K tomu mohou sloužit dvě kritéria:

- stav, ve kterém se na konci automat nachází, patří do množiny přijímajících stavů, nebo
- zásobník je na konci prázdný.

Obě definice jsou ekvivalentní, automaty na sebe lze vzájemně převádět (u druhé možnosti je možno z definice automatu zcela vypustit množinu přijímajících stavů).

Konfigurace automatu se dá popsat uspořádanou trojicí (q, w, α) , kde q je vnitřní stav, w dosud nezpracovaná část vstupu a α obsah zásobníku. Na počátku práce je automat v konfiguraci (q_0, w, z_0) .

Příklad akceptace řetězce zásobníkovým automatem: viz cvičení 10 (LL gramatiky) na courseware FJP

Vztah bezkontextových gramatik a zásobníkových automatů

Zásobníkové automaty jsou ekvivalentní bezkontextovým gramatikám: pro každou bezkontextovou gramatiku existuje zásobníkový automat, který generuje (akceptuje) identický jazyk generovaný touto gramatikou a naopak.

Pro danou BKG gramatiku $W=(N, T, P, S)$ můžeme sestrojít zásobníkový automat P takový, že $L(W)=L(P)$. Jsou dvě varianty:

1. Konstrukce zásobníkového automatu, který je modelem **syntaktické analýzy shora dolů**:
 - $Q = \{q\}$ (automat má jen jeden vnitřní stav),
 - T je shodná s množinou terminálních symbolů rozpoznávané gramatiky,
 - $G = N+T$, tj. v zásobníku se může vyskytnout jakýkoliv symbol rozpoznávané gramatiky,
 - δ je dáno rozkladovou tabulkou,
 - $q_0 = q$, počáteční stav automatu je q , neboť automat jiné stavy nemá,
 - $z_0 = S$, tj. na počátku je v zásobníku startovací symbol gramatiky

- $F = \{\}$, což se interpretuje jako "automat akceptuje vyprázdněním zásobníku".

- *LL(k) gramatiky*

2. **Analýza zdola nahoru** je obecnější a vyžaduje trochu složitější automat:

- $Q = \{q, r\}$, stav q je "pracovní", stav r "akceptační",
- T je shodná s množinou terminálních symbolů rozpoznávané gramatiky,
- G je v nejjednodušším případě rovno $N+T+\{\#\}$, tj. sjednocení symbolů gramatiky a speciálního symbolu "#"; deterministický automat může mít množinu G složitější
- δ je dáno rozkladovou tabulkou,
- $q_0 = q$,
- $z_0 = \#$,
- $F = \{r\}$.
- *gramatiky: LR, SLR, LALR*

From <<https://d.docs.live.net/e3534876709763a3/Dokumenty/ZCU/Statnice/Statnice.docx>>