

SYNTAX-DIRECTED TRANSLATION BASED ON DEEP PUSHDOWN AUTOMATA

Peter Solár

Master Degree Programme (2), FIT BUT
E-mail: xsolar05@stud.fit.vutbr.cz

Supervised by: Alexander Meduna
E-mail: meduna@fit.vutbr.cz

ABSTRACT

This paper introduces basic idea of syntax-directed translation based on deep pushdown automata. Deep pushdown automata are more powerful than pushdown automata but they are non-deterministic. We examine determinism in two ways – strict determinism and determinism with respect to the expansions. The most important model for translation is deep pushdown transducer. Also there is presented backtracking algorithm for syntax analysis, which is deterministic with respect to the depth of expansions.

1. ÚVOD

Tato práce představuje syntaxí řízený překlad založený na hlubokých zásobníkových automatech. Tento typ syntaxí řízeného překladu je, na rozdíl od běžného překladu, založen na tzv. stavových gramatikách, které jsou silnější než bezkontextové gramatiky, a tvoří nekonečnou hierarchii tříd mezi bezkontextovými a kontextovými jazyky. Další část textu je zaměřená na syntaktickou analýzu a předpokládá znalost jednotlivých fází překladu. Pro případné zopakování doporučuji studijní oporu předmětu Formální jazyky a překladače [3].

2. HLUBOKÝ ZÁSOBNÍKOVÝ AUTOMAT

Hluboké zásobníkové automaty jsou rozšířeným typem klasických zásobníkových automatů. Princip činnosti je u obou typů obdobný – provádí expanze a vyjmutí. Expanze spočívá v nahrazení neterminálního symbolu řetězcem terminálních a neterminálních symbolů. Hlavním rozdílem mezi těmito typy automatů je však to, že hluboký zásobníkový automat nekontroluje pouze symbol na vrcholu zásobníku, ale může provádět expanze i hlouběji v zásobníku (hloubka je omezena typem automatu). Tímto je dosaženo zvýšení generativní síly a tento automat je schopen generovat i některé kontextové jazyky. Pro podrobnější informace o této tématice doporučují článek prof. Meduny [2].

2.1. FORMÁLNÍ DEFINICE

Hluboký zásobníkový automat je reprezentován sedmicí ${}_nM = (Q, \Sigma, \Gamma, \delta, s, S, F)$, kde $n \in I$ je přirozené číslo označující maximální hloubku, v níž může dojít k nahrazení, Q je konečná množina stavů, Σ je konečná množina vstupních symbolů (vstupní abeceda), Γ je

konečná množina zásobníkových symbolů (zásobníková abeceda), $\Sigma \subset \Gamma$, $\#$ je speciální zásobníkový symbol značící dno zásobníku, $\# \in \Gamma - \Sigma$, $s \in Q$ je počáteční stav, $S \in \Gamma$ je počáteční zásobníkový symbol, $F \subseteq Q$ je konečná množina koncových stavů a δ je zobrazení z $I \times Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma$ do množiny konečných podmnožin množiny $Q \times \Gamma$.

2.2. DETERMINISMUS

Výše uvedená definice popisuje nedeterministickou variantu hlubokých zásobníkových automatů, která nepovoluje ε -pravidla, tj. pravidla, která dovolují neterminální symbol přepsat prázdným řetězcem. Nedeterminismus je však u překladačů, kterým se dále budeme věnovat, nežádaným jevem. Determinismus u hlubokých zásobníkových automatů můžeme chápat dvěma různými způsoby.

Prvním způsobem je striktní determinismus. Základní myšlenkou je předpoklad, že v dané konfiguraci může automat provést pouze jediný přechod (dohromady pro všechny přípustné hloubky). U této varianty není dokázáno, zda je stejně silná jako nedeterministická varianta.

Slabší formou determinismu můžeme chápat determinismus s ohledem na hloubku. Zde předpokládáme použití jednoho pravidla pro každou přípustnou hloubku. Lze dokázat, že ke každé stavové gramatice existuje deterministický zásobníkový automat s ohledem na hloubku expanzí (viz. [2]).

3. HLUBOKÝ ZÁSOBNÍKOVÝ PŘEVODNÍK

Základními modely pro jednotlivé fáze překladu jsou syntaxí řízená překladová schémata a převodníky. Vzhledem k zaměření této práce se budeme zabývat především převodníky. Hluboký zásobníkový převodník získáme rozšířením hlubokého zásobníkového automatu o možnost výstupu řetězu symbolů nad výstupní abecedou. Formálně jde o osmici ${}_nM = (Q, \Sigma_I, \Gamma, \Sigma_O, \delta, s, S, F)$, kde jednotlivé symboly mají stejný význam jako u hlubokého zásobníkového automatu, až na $\Sigma_I = \Sigma$, $\Sigma_I, \Sigma_O \subseteq \Sigma$, $\# \in \Gamma - \Sigma_I - \Sigma_O$, δ je zobrazení z $I \times Q \times (\Sigma_I \cup \{\varepsilon\}) \times \Gamma$ do množiny konečných podmnožin množiny $Q \times \Gamma \times \Sigma_O^*$.

3.1. DETERMINISMUS

Hluboký zásobníkový převodník má stejně definovány typy determinismu jako hluboký zásobníkový automat.

Hluboký zásobníkový převodník ${}_nM = (Q, \Sigma_I, \Gamma, \Sigma_O, \delta, s, S, F)$ nazýváme deterministickým s ohledem na hloubku expanzí, pokud splňuje následující podmínky:

- 1, $\forall q \in Q, \forall a \in \Sigma_I \cup \{\varepsilon\}, \forall B \in \Gamma$: mohutnost množiny $\delta(i, q, a, B) \leq 1$
- 2, $\forall q \in Q, \forall a \in \Sigma_I, \forall B \in \Gamma$: jestliže $\delta(i, q, \varepsilon, B) = \emptyset$, pak $\delta(i, q, a, B) = \emptyset$

3.2. KONFIGURACE

Konfigurace hlubokého zásobníkového převodníku je čtveřice (q, x, z, y) , kde $q \in Q$ je aktuální stav převodníku ${}_nM$, $x \in \Sigma_I^*$ je dosud nepřečtená část vstupního řetězce, $z \in \Gamma^*$ je obsah zásobníku a $y \in \Sigma_O^*$ je dosud vytvořený část výstupního řetězce.

Analogicky k hlubokým zásobníkovým automatům (viz. [2]) můžeme i u hlubokých zásobníkových převodníků zavést relaci přechodu mezi jednotlivými konfiguracemi \Rightarrow .

4. ALGORITMUS

Navržená metoda syntaxí řízeného překladu založeného na hlubokých zásobníkových automatech je inspirována obecnou syntaktickou analýzou a při své práci využívá metodu backtracking (tj. využívá další zásobník pro historii).

```
correct_syntax := false;
do {if (na vrcholu zásobníku je terminál)
    {if (vstup je prázdný) {obnovení konfigurace systému ze zásobníku historie}
    else {if (na vrcholu zásobníku je symbol shodný se vstupním symbolem)
        {vyjmutí symbolu z vrcholu zásobníku, posun čtecí hlavu na další vstupní symbol a poznačení expanze na zásobník historie}
        else {obnovení konfigurace systému ze zásobníku historie}}}
    else { for (i = 1, i <= n) // n = maximální hloubka
        { r = nalezni pravidlo hloubky i s daným neterminálem a aktuálním vstupním symbolem;
        if (pravidlo r nalezeno) break;
        }
        if (pravidlo r nalezeno) { uložení aktuálního stavu na zásobník historie a následná aplikace nalezeného pravidla r }
        else obnovení konfigurace systému ze zásobníku historie}
    }
    if (zásobník je prázdný) correct_syntax := true;
} while ((zásobník historie není prázdný) and (correct_syntax == false));
```

5. ZÁVĚR

V této práci je prezentována hlavní myšlenka syntaxí řízeného překladu založeného na hlubokých zásobníkových automatech. Základním modelem části syntaktické analýzy je definován hluboký zásobníkový převodník, jako silnější alternativa k zásobníkovým převodníkům užívaným u klasické syntaktické analýzy. Nejdůležitější částí tohoto příspěvku bylo uvedení algoritmu, podle něž se provádí samotná syntaktická analýza. Zbylé části překladu budou detailně rozebrány v mé diplomové práci.

LITERATURA

- [1] Meduna, A.: Automata and Languages, London, Sprinter, 2000
- [2] Meduna, A.: Deep pushdown automata, Acta Informatica, 98, 2006, s. 114-124
- [3] Meduna, A., Lukáš. R.: Formální jazyky a překladače, studijní opora, Brno, FIT VUT v Brně, verze 1.2006