

SYNTAX ANALYSIS BASED ON COMBINATION OF SEVERAL METHODS

Martin ČERMÁK, Bachelor Degree Programme (3)
Dept. of Computer Systems, FIT, BUT
E-mail: xcerma16@stud.fit.vutbr.cz

Supervised by: Prof. Alexander Meduna

ABSTRACT

This paper introduces and discusses *syntax analysis based on combination of several methods* as a new method of the syntax analysis. This new method is based on generation of a derivation tree with the aid of more than one original methods of syntax analysis, where each of them generates part of the tree. The present paper discusses sequential and parallel generation and their usability in a practice. In its conclusion, this paper suggest combination of top-down and bottom-up parsing as a good solution for the syntax analysis.

1 ÚVOD

Syntaktický analyzátor je nejvýznamnější část překladače, která má za úkol zanalyzovat text (za pomoci lexikálního analyzátoru), napsaný v daném jazyce, a vytvořit pro něj abstraktní syntaktický strom. Ten nám říká, jakým způsobem byl text zpracováván a zda vůbec překládaný text do příslušného jazyka patří. Abstraktní syntaktický strom je vytvářen na základě množiny gramatických pravidel určujících použitelné větné formy pro konkrétní jazyk.

2 KOMBINACE METOD

Pro syntaktickou analýzu (SA) existuje řada metod, které můžeme rozdělit podle přístupu do dvou základních kategorií. Jedná se o SA shora dolů a SA zdola nahoru. Každá z metod, reprezentující kteroukoliv z těchto kategorií, je pak popisována jednou příslušnou gramatikou. Problém nastává v okamžiku, kdy začneme hledat metodu SA, která by byla nejvhodnější pro popis daného jazyka. Při bližším zkoumání zjistíme, že některé části zpracovávaného jazyka by bylo vhodnější popisovat jednou metodou a jiné části zase metodou jinou. Kombinováním metod můžeme získat zcela nové parametry syntaktického analyzátoru.

2.1 PODMÍNKY KOMBINOVÁNÍ METOD

Aby správně kombinování metod pracovalo, je nutné zvolit jednu „vůdčí“ metodu, která

bude celý běh překladu řídit. Dále je nutné specifikovat způsob, jakým budou jednotlivé metody spolu navzájem komunikovat, a způsob jakým bude přechod na jinou metodu vůbec realizován.

3 PŘEPÍNANÁ SYNTAKTICKÁ ANALÝZA (PSA)

Přepínanou SA rozumíme kombinovanou SA využívající přepínacích pravidel a aparátu pro přepínanou SA.

3.1 PŘEPÍNACÍ PRAVIDLO

Necht' $A \rightarrow B$ je pravidlem z gramatiky G ,

kde:

- $A, B \in (N+T)^*$,
- N je množinou neterminálních symbolů,
- T je množinou terminálních symbolů.

Toto pravidlo nazveme přepínacím pravidlem, pokud je jeho užitím realizován přechod mezi metodami syntaktické analýzy.

3.2 APARÁT PRO PŘEPÍNANOU SYNTAKTICKOU ANALÝZU (APSA)

Aparát pro přepínanou syntaktickou analýzu je osmice $APSA=(G, N, T, M, \Gamma, \Sigma, \Psi, \Lambda)$, kde:

- G je množina gramatik,
- N je množina množin neterminálních symbolů,
- T je množina terminálních symbolů,
- M je množina zásobníkových automatů,
- Γ je množina množin zásobníkových symbolů,
- Σ je vstupní abeceda,
- Ψ je množina množin přepínacích pravidel,
- Λ je množina metod syntaktické analýzy.
- $|\Lambda| \geq |\Psi|$
- $|\Psi|, |G|, |M| \geq 1$

3.3 PARALELNÍ PŘEKLAD

Pomocí definic z předchozích dvou kapitol jsme dostali metodu SA, jejíž parametry můžeme libovolně měnit v závislosti na zvolených metodách a výběru jejich podílu práce na konkrétním zpracování textu daného jazyka. Vliv na parametry syntaktického analyzátoru má do značné míry i zvolení vhodného počtu zásobníkových automatů. Metody, díky přepínacímu pravidlu, mohou pracovat i na jednom společném zásobníku. Pokud ale zvolíme větší počet zásobníků, můžeme bez větších problémů efektivně řešit paralelní překlad.

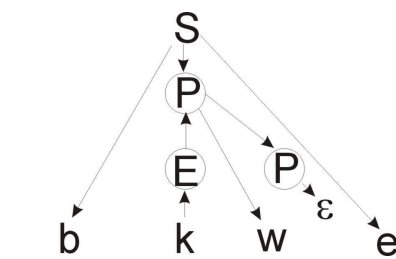
4 PREDIKTIVNÍ SA PŘEPÍNANÁ NA LR-SA

Z vlastností metod syntaktických analýz se ukázalo, že je velmi výhodné kombinovat metody pro SA shora dolů a zdola nahoru. LL-SA se špatně potýká s levě rekurzivními pravidly. Oproti tomu SA zdola nahoru je buď příliš slabá na zpracování celého zdrojového kódu, nebo naopak může velmi výrazně ztížit návrh SA. Je tedy vhodné použít LL-SA pro řízení toku SA a některou z metod SA zdola nahoru pro zpracovávání výrazů.

4.1 PŘÍKLAD APSA S VYUŽITÍM JEDNOHO EPDA

$G=(G_1, G_2)$, $G_1=(1:S \rightarrow bPe, 2:P \rightarrow wP, 3:P \rightarrow \epsilon)$, $G_2=(1:E \rightarrow EoE, 2:E \rightarrow k, 3:wE \rightarrow w)$. Tyto gramatiky generují jazyk $L=w(k(ok)^*w)^*e$. Postup syntaktické analýzy pro vstupní řetězec „bkwe“:

1:	Vstup: bkwe, Zásobník: [S,	pravidlo pro další krok: G1:1
2:	Vstup: bkwe, Zásobník: [ePb,	pravidlo pro další krok:
3:	Vstup: kwe, Zásobník: [eP,	pravidlo pro další krok: G1:2
4:	Vstup: kwe, Zásobník: [ePw,	pravidlo pro další krok:
5:	Vstup: we, Zásobník: [ePwk,	pravidlo pro další krok: G2:2
6:	Vstup: we, Zásobník: [ePwE,	pravidlo pro další krok: G2:3
7:	Vstup: we, Zásobník: [ePw,	pravidlo pro další krok:
8:	Vstup: e, Zásobník: [eP,	pravidlo pro další krok: G1: 3
9:	Vstup: e, Zásobník: [e,	pravidlo pro další krok:
10:	Vstup: , Zásobník: [,	pravidlo pro další krok: Konec



Obr. 1: Derivační strom

4.2 SOUHRN

Z obrázku je patrné, že bude výsledná SA o něco málo pomalejší než SA založená čistě na LR. Některé části překladu lze řešit i paralelně, což překlad opět zrychlí.

LITERATURA

- [1] Češka, M., Hruška, T., Beneš, M.: Překladače – Skriptum VUT Brno. Ediční středisko VUT Brno.