

Oponentský posudok na doktorskú dizertačnú prácu
RNDr. Michala Žemličky
"Principles of Kind Parsing"

Dizertačná práca je venovaná skúmaniu syntaktickej analýzy v bezkontextových gramatikách (presnejšie, určitých variantov týchto gramatík), ako aj problematike parsingu v prekladačoch pre takéto gramatiky, metódou typu „top-down“, a niektorými jej modifikáciami. Okrem iného je tu skúmaný aj problém modifikácie syntaxe jazyka popísaného gramatikou užívateľom počas parsingu vstupného textu.

Aktuálnosť riešenej problematiky a aplikácia výsledkov v praxi: Aj keď v súčasnosti nepatrí uvedená problematika medzi prioritné smery výskumu v oblasti softwarových systémov, keďže bola veľmi detailne skúmaná už minulosti, stále ju možno považovať za dôležitú, keďže prekladače sú nevyhnutným nástrojom pre tvorbu nových programových aplikácií. Navyše sa parsing zjavuje a využíva i v oblastiach, kde v minulosti využívaný nebol, ako je napríklad tvorba dokumentov v systémoch typu TeX a jeho nadstavbách, či pri aplikáciách na báze XML. TeX je navyše typickou aplikáciou, kde sa masovo využíva jeden z fenoménov skúmaných v dizertácii, jedná sa tu totiž o systém, ktorý počas parsingu dynamicky rozširuje svoju syntax.

Výsledky a ich význam pre ďalší rozvoj vedného odboru:

(A) Ťažiskom práce je zavedenie nového typu CF gramatík (t.j. bezkontextových gramatík), takzvaných „kind grammars“ (príjemných gramatík), symbolicky označovaných ako $K(k)$ gramatiky. Motivácia pre zavedenie $K(k)$ gramatík je nasledovná: Jeden z najznámejších a najčastejšie používaných typov CF gramatík sú $LL(k)$ gramatiky, pre ktoré sa dá „pohodlným a jednoduchým“ spôsobom konštruovať deterministický parser pracujúci metódou zhora nadol. Nevýhodou týchto gramatík je ale to, že neumožňujú pri popise syntaxe jazyka používať ľavú rekúziu, t.j. odvodenia typu $A \Rightarrow \dots \Rightarrow Au$, pretože v parseri procedúra pre neterminál A by ako prvú akciu volala procedúru pre neterminál A , čím by došlo k nekonečnému plneniu zásobníka. Na druhej strane, ľavá rekúzia je veľmi prirodzeným vyjadrovacím prostriedkom pri definícii syntaxe jazyka. Je síce známe, že ľavú rekúziu možno z gramatiky odstrániť, napríklad jej transformáciou na Greibachovej normálnej tvar, ale pri takýchto transformáciách sa stráca priamy súvis so sémantickými akciami asociovanými ku jednotlivým gramatickým pravidlám.

Preto autor zavádza uvedené $K(k)$ gramatiky, ktoré zachovávajú „dobré“ vlastnosti $LL(k)$ gramatík, ako napr. deterministický top-down parsing, no $K(k)$ gramatiky navyše umožňujú tzv. priamu ľavú rekúziu, t.j. pravidlá typu $A \Rightarrow Au$. Nepriama a skrytá ľavá rekúzia, t.j. odvodenia typu $A \Rightarrow \dots \Rightarrow Au$ pomocou viacerých krokov sú ale naďalej obmedzené. (Na rozdiel od priamej sú ale nepriama a skrytá ľavá rekúzia pri definícii jazyka v praxi raritou.) Jemným rozšírením na tzv. slabý variant sa potom zavádza pojem $WK(k)$ gramatiky („weak kind grammar“), ktoré sú nadmnožinou jednak $K(k)$, ako aj $LL(k)$ gramatík, pre každé $k \geq 1$. (Pre prax najdôležitejším prípadom je $k=1$, kde z hľadiska generatívnej sily platí $WK(1)=K(1) > LL(1)=SLL(1)$.) Formálnym dôkazom, ktoré demonštrujú generatívnu silu $K(k)$ a $WK(k)$ gramatík a zaraďujú $K(k)$ a $WK(k)$ gramatiky do hierarchie medzi známe typy CF gramatík, ako sú $LL(k)$ a $SLL(k)$ gramatiky, je venovaná kapitola 6. Aplikácie a popis deterministického top-down parsingu pre tento typ gramatík je popísaný v kapitolách 8-11.

(B) Podobná motivácia, ale v menšom rozsahu, viedla ku skúmaniu tzv. „lookahead“ zásobníkových automatov. Hlavným cieľom bolo dosiahnuť nízku popisnú zložitosť zásobníkového automatu (t.j. malý počet stavov a jednoduchú prechodovú funkciu), a to tým, že

zásobníkový automat vybavíme dodatočnou schopnosťou preskúmať k symbolov dopredu pred čítacou hlavou, napríklad pomocným konečným automatom, alebo, v obecnom prípade, použitím iného typu „orákula“.

Mám pripomienku ku formálnemu zavedeniu tzv. „extensible grammars“, t.j. gramatík, ktoré sú schopné rozširovať svoju syntax. V práci sa síce formálny popis takéhoto mechanizmu nachádza (kapitola 9.4), ale mám pocit, že v prípade, ak má dochádzať ku rozširovaniu syntaxe počas spracovávania vstupu (syntax si rozširuje užívateľ „on-line“), je problém tak trochu „schovaný pod posteľ“, t.j., modifikácia syntaxe je ponechaná na sémantickú úroveň. Nenašiel by sa nejaký vhodnejší mechanizmus, ktorý by rozširovanie syntaxe dokázal vybaviť na syntaktickej úrovni?

Kvalita formálneho spracovania dizertácie: Práca je napísaná v angličtine, jasne a zrozumiteľne. Autor predkladanou prácou dokazuje, že zvládol rôzne dôkazové techniky teórie formálnych jazykov, čo sa týka najmä kvality kapitol 5, 6, a 7, ktoré obsahujú nové výsledky teoretického charakteru. Na druhej strane, kapitoly 4 a 12 sú výslovne na úrovni populárnovedeckej. Kapitola 2 má názov „notation and known results“ (označenia a známe výsledky), v skutočnosti však obsahuje iba základné definície, plus znenie dvoch viet z klasickej literatúry, ktoré sú síce neskôr v práci potrebné, ale nesúvisia s doterajším stavom problematiky. Pri prvom čítaní má tak čitateľ na začiatku dojem, že autor nevenoval pozornosť prehľadu doteraz dosiahnutých výsledkov známych z literatúry. Tie sú uvedené dosť netradične, na záver, v kapitole 13.

Záverom: Práca prináša nové poznatky a preukazuje spôsobilosť RNDr. Michala Žemličky ku samostatnej a tvorivej vedeckej práci. Preto doporučujem, aby na základe predloženej dizertácie, berúc do úvahy aj fakt, že výsledky boli okrem iného prezentované v zborníkoch troch renomovaných medzinárodných konferencií, bola jej autorovi udelená vedecká hodnosť Ph.D.

V Košiciach, 23.októbra 2006



Prof. Dr. habil. Genert, DrSc.,
Ústav Informatiky Prírodovedeckej fakulty
Univ. P.J.Šafárika v Košiciach