

Posudek vedoucího doktorské disertační práce

*RNDr. Michal Žemlicka*  
**Principles of Kind Parsing.**

Cílem práce bylo nalezení gramatik, které si zachovávají základní softwarově inženýrské výhody LL gramatik a při tom nemusí být úplně faktorizované zleva. Tento cíl byl splněn.

Autor našel třídy gramatik (tzv. přívětivé gramatiky) s výhledem  $k$ , o nichž dokázal, že do značné míry splňují výše uvedené požadavky, což prokázal i implementací rozšiřitelného parseru pro přívětivé gramatiky, který připouští velkou volnost volby míst, kde je možné umístit metadefinice. Podobné vlastnosti má i autorem implementovaný konstruktor pro  $k=1$ . Výstupem implementovaného konstruktoru je program, který je ve tvaru vhodném pro případná ruční doplnění. To je mnohdy žádoucí, např. při aplikacích v lingvistice (reakce na nové jevy, vyšší formy interakce s uživateli) a optimalizaci testů výhledu.

Motivací byl problém gramatiky pro aritmetické výrazy a pro gramatiky některých dalších konstrukcí programovacích jazyků (tyto gramatiky jsou levě rekurzivní) a snaha připustit i pravidla, které mohou mít společný prefix. To umožňuje více používat v parserech pravidla ve tvaru, jak je definoval lingvista nebo navrhovatel jazyka či jak syntax jazyka chápou jeho uživatelé. To je výhodné při spolupráci s neinformatiky při návrhu a implementaci parserů, např. pro převod dokumentů do formátu XML. Blízkost kind gramatik LL gramatikám umožňuje softwarově inženýrsky výhodné varianty implementace a využívání parserů a konstruktorů. To může být využito např. při výše zmíněné transformaci dokumentů do XML formy. Jedná se tedy o výsledky s potenciálem významného využití v praxi.

Autor dokázal, že přívětivé gramatiky s výhledem  $k$  generují právě  $LL(k)$  jazyky. Intuitivně, jestliže použijeme při konstrukci parseru grafovou reprezentaci množin pravidel tak, jak ukazuje autor, to není překvapující. Formální důkaz je ale poměrně zdlouhavý. Vyžaduje totiž konstrukci a důkaz správnosti gramatik potřebných pro důkazy známých tvrzení o levé faktorizaci bezkontextových gramatik. Tyto gramatiky se nepodařilo přes značné úsilí v dosažitelné literatuře nalézt a musely být zkonstruovány, což platí i pro podrobné důkazy o levé faktorizaci.

Autorova implementace parseru i konstruktoru usnadňuje různé metody optimalizace vyhodnocování výhledu. Některé varianty vyhodnocování výhledu jsou v práci podrobněji diskutovány. Na příkladech je ukázáno, že tato optimalizace je výhodná už v případě  $k=1$ .

Hlavní výhodou takového řešení je ale fakt, že proces vyhodnocování výhledu je možné chápat (a implementovat) jako orákulum (černou skříňku) vracející informaci, kterou akci je nutné/možné v určité konfiguraci provést. Orákulum může využívat konečný automat, vyhodnocující potenciálně nekonečný výhled, může ale generovat dotazy na uživatele nebo komunikovat s jinými aplikacemi. V tom případě jde o formalizaci vhodnou pro situace, kdy je žádoucí uplatnit servisně orientovanou architekturu (SOA, to je např. případ doplňování uživatelských variant pohledu na strukturu přirozeného jazyka do dat parseru).

Výsledky disertace byly publikovány v několika sbornících z mezinárodních konferencí (z toho jedna v LNCS, další vyjde v LNCS letos), jedné kapitole knihy (vyšla v Londýně) a v jednom článku v mezinárodním časopise. Výsledky související s disertační prací tvoří jen menší část publikačních aktivit autora. Autor je úspěšný i jako konsultant doktoranda (3 mezinárodní publikace). Autor tedy prokázal schopnost samostatné vědecké práce.

Domnívám se, že práce splňuje podmínky kladené na doktorské disertační práce a práci vřele doporučuji k obhajobě.

V Praze dne

