

Oponentský posudek dizertační práce

Název práce: Synchronous formal systems based on grammars and transducers

Autor: Ing. Petr Horáček

Školitel: prof. RNDr. Alexander Meduna, CSc

Posudek

Shrnutí výsledků

Předkládaná dizertační práce Ing. Petra Horáčka je prací z oblasti teorie formálních jazyků, automatů a gramatik, která tvoří jeden ze základů teoretické informatiky. Práce obsahuje jak řešeršní části existujících výsledků, tak i části, ve kterých autor přináší vlastní nové poznatky. V práci je výborným způsobem shrnuta základní teorie principů převodníků, synchronních gramatik a řízeného přepisování, také s ohledem na použití v oblasti výpočetní lingvistiky. Práce zavádí nové formální systémy a zkoumá jejich vlastnosti, konkrétně se jedná o pravidly synchronizované gramatiky (rule-synchronized context-free grammars, synchronous scattered context grammars a synchronous matrix grammars) a pravidly omezené převodníky (rule-restricted transducers).

Synchronizované gramatiky jsou založeny na myšlence přidání druhé gramatiky, u které je použití pravidel synchronizováno s použitím pravidel první gramatiky (a naopak) pomocí tzv. synchronizačních funkcí.

Vedle zavedení těchto synchronních gramatik se nejvýznamnějším přínosem práce ukazují být nové teoretické výsledky týkající se jejich popisné generativní síly. Tato část práce obsahuje výsledky podrobného teoretického výzkumu, které jsou velice srozumitelně prezentovány.

Dále práce zkoumá generativní a přijímací sílu několika variant převodníku založeného na kombinaci konečného automatu a bezkontextové gramatiky.

Z praktičtějšího pohledu práce obsahuje rozbor perspektiv aplikací navržených formálních modelů ve zpracování přirozeného jazyka se zaměřením na překladové systémy. Na jednotlivých případových studiích ukazuje, jak lze zkoumané systémy využít pro formální popis syntaktických struktur v anglickém, českém a japonském jazyce a pro popis vzájemných transformací odpovídajících struktur mezi těmito jazyky.

Formální úroveň práce

Po formální stránce je práce napsána na výborné úrovni - práce neobsahuje formální ani věcné chyby, po jazykové stránce je psaná srozumitelnou angličtinou bez chyb.

Budoucí práce

Co se týká budoucí práce, pak konstatuji, že v teoretické rovině lze dále studovat vlastnosti navržených formálních modelů. Práce například ukazuje, jaké třídy jazyků modely definují, ale není zatím známo, jak efektivně (tzn. jak lze omezit např. počet symbolů nebo pravidel). Je zde tedy možnost studovat popisnou složitost modelů. Také lze zkoumat další omezení na tyto formální systémy (např. bez vymazávacích pravidel, s různým omezením nedeterminismu) a obecně další varianty systémů. Z praktického pohledu je třeba dále studovat možnosti začlenění navržených modelů do komplexnějších překladových systémů. Dalším otevřeným problémem zůstává efektivita syntaktické analýzy, především při použití modelů s řízeným přepisováním. Bylo by potřeba zpracovat algoritmy převodu synchronních gramatik na příslušné typy automatů.

Výhrady k práci

Z výhrad k práci bych uvedl následující hlavní výhrady: Prakticky orientované aspekty práce by mohly být rozpracovány podrobněji, což by zesílilo motivační důvody pro zavádění nových formalismů do praktického používání. V současné podobě práce obsahuje z tohoto pohledu pouze konkrétní případové studie a obecné srovnání výhod navržených formálních modelů. Část věnovaná aplikacím by zejména mohla zahrnovat experimentální implementaci překladového systému založeného na navržených modelech a její následné testování.

V práci by také mohlo být detailnější srovnání navrhovaných postupů s existujícími metodami pro formální popis přirozených jazyků a s existujícími překladovými systémy.

Dotazy na autora práce k obhajobě

1. Každá ze zaváděných synchronních gramatik spojuje vždy dvě gramatiky přesně stejného typu (CFG a CFG, SCG a SCG, MAT a MAT). Uvažoval jste nad analogickými spojeními dvou gramatik, kde by tyto dvě gramatiky nebyly stejného typu a synchronizace pomocí druhé gramatiky mohla tak být podobná (např. CFG a CFG, MAT a GFG)?
2. Nejjednodušším zaváděným případem synchronních gramatik je synchronní bezkontextová gramatika. Proč se autor nezabýval také analogickými synchronními regulárními gramatikami?
3. Co by znamenalo pro generativní síly synchronních gramatik zavedení omezení synchronních funkcí v gramatikách tak, aby zobrazení pro používání pravidel z jedné množiny pravidel do druhé množiny pravidel bylo např. injektivní?
4. Synchronní gramatiky jsou definovány jako spojení dvou gramatik. Co by z teoretického pohledu přineslo analogické rozšíření na spojení obecně $n \geq 2$ gramatik.

Celkové hodnocení

Celkově konstatuji na základě předložené práce, že Ing. Petr Horáček jednoznačně prokázal své znalosti a schopnosti potřebné pro úspěšnou vědeckou práci. Proto práci velmi rád **doporučuji k obhájení** za účelem získání titulu Ph.D.

V Utrechtu 18.7.2014,

doc. Ing. Jan Janoušek, Ph.D.

vedoucí katedry teoretické informatiky,

Fakulta informačních technologií,

České vysoké učení technické v Praze