

Posudok na dizertačnú prácu

Mgr. Martin Procházka: Redukční automaty a syntaktické chyby.

Predkladaná práca sa venuje redukčným automatom v súvislosti s identifikáciou špecifických syntaktických chýb. Zavádza pojem normovaného redukčného automatu, čo zahŕňa determinizmus, monotónnosť, prefixovú korektnosť a minimálny počet stavov. Pre tento typ redukčného automatu viacmenej priamočiaro konštruuje tzv. redukčný analyzátor, ktorý vkladáním špeciálnych značiek identifikuje v spracovávanom slove (ne)korektné úseky. Formálnym dôkazom vlastností analyzátoru dokazuje korektnosť konštrukcie. Požiadavka na minimalitu počtu stavov sa v kontexte tejto práce zdá byť umelá. A to napriek tomu, že minimalizácia počtu stavov má zmysel. V práci je podaná netriviálna konštrukcia normalizácie deterministického monotónneho red-automatu, ktorá prakticky rozširuje využiteľnosť navrhnutého analyzátoru.

Práca je príspevkom k uchopeniu reálnych problémov oblasti lingvistiky modelmi a metódami teórie automatov. Je to aktuálny smer, ktorý je akýmsi prepojením teórie a praxe.

Úvod práce predznačuje spôsob písania práce. Sústreduje sa na to, čomu sa práca venuje, nemotivuje, nevyясňuje prečo. Formuláciu cieľa práce v úvode nenájdeme, cieľ treba hľadať roztrúsený v ďalších kapitolách. Je to škoda, komplikuje to hodnotenie práce. Očakávam formuláciu cieľa a zhodnotenie vlastného prínosu na obhajobe.

Prvá kapitola je venovaná definícii nového modelu, ktorým je redukčný automat. Ide o deterministický reštartovací automat, ktorý má povolené modifikovať spracovávané slovo tak, že tesne pred reštartom vynechá v sufixe ohraničenej veľkosti niektoré znaky. Autor sa snaží motivovať definovanie modelu poukazovaním na rozdiely definovaného automatu a reštartovacieho automatu; s jeho argumentáciou nesúhlasím (a sám si ju neskôr tiež vyracia). Podľa môjho (a aj autorovho) názoru sú Red-A špeciálnym prípadom RA - sú skracovacie a prepis vzniká vynechávaním. Presumutie časti, ktorá sa prepisuje, z okna do stavu a následná zmena smerom dočasne pred snímaným symbolom je len otázkou pohľadu a inou realizáciou okna reštartovacieho automatu. Analyzátor sa následne vracia k používaniu pomocných symbolov; je teda tiež špeciálnym prípadom reštartoviacích automatov, keď sa upravuje, akým spôsobom sa môžu používať prepisovacie symboly a z lingvistickejho hľadiska zmysluplnie interpretuje ich význam. Definovanie nového typu, skor jeho pomenovanie, je v situácii, keď sa ním ďalej zaoberáme a rôznych typov reštartoviacích automatov je veľa, zmysluplné. Je možné, že práve identifikovanie tohto modelu je jedným z (nevypovedaných) cieľov práce. Pár slov o vzťahu definovanej triedy jazykov k iným - aplikačným triedam jazykov by však bolo na mieste.

Druhá kapitola sa venuje robustnej redukčnej analýze chybnych slov k jazykom zadaným normovaným monotónnym Red-automatom. Až tu sa dozvedáme, že cieľom je lokalizácia podľa možnosti všetkých skutočných syntaktických chýb. Lokalizuje sa

→ maximálny korektný prefix (Treba povedať, že som mala problém so zavedením pojmu korektného jadra. Korektné jadro slova w (vzhľadom k jazyku L) je také jeho podľa, ktoré sa dá "doplniť" prefixom a sufíxom na slovo z L a pritom je maximálne také. Trvalo mi dlho, kým som si uvedomila, že maximálne znamená, že ono samé sa už v rámci slova w nedá predĺžiť na korektné podľa.) a

→ maximálna postprefixová (ne)korektnosť, teda maximálny sufíx korektného jadra.
Za nedostatok považujem, že *na tomto mieste* chýba diskusia o tom, prečo hľadáme práve tie. V úvodnej kapitole sa zmieňuje, že to vychádza z neopravujúceho zotavenia zo syntaktických chýb a Richterovho konceptu maximálnych súvislých korektných podľa a minimálnych súvislých nekorektných podľa. V čom sa líši autorov prístup?

Nie je šťastné, že namiesto vysvetlenia sa na čitateľa valia formálne definície pojmov a sám musí hľadať, čo bolo dovodom pre práve takéto definície (aj keď niekedy sa to vysvetlenie objaví neskôr). Prečo sa hľadajú najdlšie sufíxy a ignorujú potenciálne skoršie chýby? Prečo sa pri identifikácii ďalšej (ne)korektnosti ignoruje doteraz spracovaná časť slova? Prečo analyzátor predpokladá, že potom, ako bol automat korektnie v stave p bude po výskytu chýby v stave q , keď stav q môže byť zo stavu p nedosiahnuteľný? ...

Nasledujúci podrobnejší popis "metódy" je korektnou realizáciou rozhodnutia, že po objavení sa chýby automat pokračuje paralelne zo všetkých nekoncových stavov. Vyžadovaná minimalita red-automatu sa v procese redukčnej analýzy nevyužíva (Veta 2.4.2 platí aj pre prefixovo korektné mon-red automaty). Načo teda minimalita počtu stavov v definícii normalizovaného automatu? V čom je (ako sa avizuje) výhodná?

Dôkaz korektnosti popísanej realizácie identifikácie chýb je formálny, doplnený príkladmi. Plusom je, že ním dizertant preukazuje schopnosť zachytiť to, čo je pre dôkaz dôležité; negatívom je, že čitateľ sám musí prieť na to, čo treba dokázať a hľadať to v teste, ktorý nasleduje.

Časť 2.5.2 začína otázkou, ktorá ma potešila: "Jaké musí byť vlastnosti mon-red automatu M , aby jeho postprefixový robustní analyzátor značku "?" nevložil do žiadneho slova?". Pozitívom je identifikovanie jednoznačnej redukovateľnosti. Nepotešilo ma, že ostalo len pri postačujúcej podmienke; ani úvaha o nutnosti, ani úvaha o tom, či je problém jednoznačnej redukovateľnosti rozhodnuteľný...

Musela som sa použiať pri čítaní (426) predpokladu o tom, že návrat hlavy je v konštantnom čase, keďže sa zložitosť vyjadruje asymptoticky. V súvislosti s diskutovaním implementácie vypuklejšie vyvstáva otázka odôvodnenia, prečo sa redukčný analyzátor od začiatku nevytváral ako program.

Tretia kapitola sa venuje normalizácii mon-Red-automatu, čím zvyšuje význam prezentovaného analyzátoru (len to dizertant nezdôrazňuje). Pomerne komplikované (konštrukciou gramatiky k automatu a následnou konštrukciou automatu k nej) sa dosahuje prefixová korektnosť mon-red-automatu, ktorá je pre konštrukciu analyzátoru dôležitá.

Výska grafickej reprezentácie red-automatu a následná konštrukcia ekvivalentnej LR(0) gramatiky sú podľa môjho názoru najobtiažnejšie konštrukcie práce. Hoci je prezentácia nie veľmi šťastná, vyžadovala kus invencie.

Myslím, že v tvrdení vety 3.1.1. chýba v znení predpoklad o neexistencii $T_0 \subset R(T_1)$, resp. informácia o tom, že p je počet maximálnych neterminálových cest.

V časti o minimalizácii monotónnych red-automatov som si na okraj poznamenala "...asi už nemá čas...". Príklady prestali, dôkazy sa až príliš skrátili.

V **závere** sa píše, že jedným zo zámerov bolo teoreticky podložiť prakticky používanú metódu panického módu. Čo je panický mód práca neuvádza a ako predložená práca k naplneniu naznačeného cieľa prispela, som sa nedozvedela.

Prílohou práce sú dva články, ktorých zaradeniu do práce uerozumiem. Prvý z nich sa v práci spomína ako anglická verzia toho, čo je v práci česky, a preto som ho nečítala. Druhý sa v závere jednou vetou zmieňuje ako možné nové vnímanie syntaktických chýb. Žiadna zmienka o obsahu, význame, výsledkoch.... Rozumiem tomu tak, že ho nemám brať za súčasť textu, na ktorý písem posudok. Nerozumiem, prečo výsledky neboli zaradené do práce v rámci samostatnej krátkej kapitolky, nakoľko obsahujú dolné odhady, ktoré by zvýšili kvalitu práce. (Prílohu tvorí neporiadne okopírovaná dosť zle napísaná verzia textu akceptovaného na workshop NCMAA 2012. Ide o takmer nekomentovaný zoznam definícií, tvrdení a dokazov. Na rozdiel od práce však tento príspevok obsahuje aj dolné odhady.) Tiež sa mi z poznámok v úvode zdá, že niektoré z prác, ktorých je doktorand autorom alebo spoluautorom, by si zaslúžili v práci väčšiu pozornosť.

Zhrnutie

Dizertant v práci používa štandardné metódy formálnych jazykov, automatov a prekladačov, ktoré zvládol a dokázal tvoriť aplikovať. Preukázal to definovaním modelu s dobre zvolenými vlastnosťami, konštrukciou analyzátoru k nemu s formálnym dokazom jeho korektnosti. Transformácia na prefixovo korektný automat je realizovaná netriviálou simuláciou red-automatu LR(0) gramatikou a následnou konštrukciou požadovaného automatu k nej. Minimalizácia počtu stavov viacmenej odkazuje na existujúce simulácie, dôkazy v tejto časti sú neprimierane krátke, nie sú sú istá ich korektnosťou.

Za výrazný nedostatok práce považujem jej prezentáciu. Práca je napísaná dostatočne formálne, prekľupy som nevidela, nemožno jej vyčítať ani absenciu detailne vypracovaných dobre zvolených príkladov; dizertant si pri písaní vhodne pomáha definovaním pojmov a uvádzaním príkladov. Je preto až neuveriteľné, že čítanie výsledných dôkazov je tak fažké; kde človek okamžite nevie spraviť dôkaz sám, je to čítanie v lepšom prípade dvojprechodom. Prezentáciu výsledkov výrazne chýba nadľhad, absentuje neformálne z nadľhadu podané vysvetlenie ideí, konštrukcií,... (Konštrukcia LR(0) gramatiky je formálne korektná, detailná, ale vysvetľovaná syntakticky, nie sémanticky.)

Ďalšie otázky pre doktoranda

- Vysvetlite, prečo je z lingvistickeho hľadiska zaujímavé identifikovanie (ne)korektných úsekov v teste tak, ako ste to robili.

- Je definovanie analyzovaného jazyka deterministickým monotónnym red-automatom vo Vami predpokladanej aplikáčnej oblasti prirodzené?

Záverom: Podstatná časť pripomienok smeruje k prezentácii, čo je zrejme spôsobené nedostatkom času pri písaní a malej skúsenosti s vysvetľovaním. Popri tom dizertant predloženou prácou a publikovanými článkami dostatočne preukázal schopnosť samostatnej tvorivej vedeckej práce.

Odporučam preto priať predkladanú prácu *Redukční automaty a syntaktické chyby* k obhajobe a doporučujem *priznať* doktorandovi mgr. Martinovi Procházkovi vedecko-akademickú hodnosť PhD.

V Bratislave, 6.8.2012

doc. RNDr. Dana Pardubská, CSc.