

Oponentský posudek diplomové práce

Jan Ptáček: Generování vět z tektogramatických stromů Pražského závislostního korpusu

Diplomová práce Jan Ptáčka byla vytvořena na základě zadání ÚFAL MFF UK jako součást experimentů v oblasti strojového překladu.

Práce je rozdělena do čtyř kapitol a dvou příloh. V první kapitole je popsán cíl práce, a rovněž motivace a datová základna, nad kterou se práce realizovala. Ve druhé kapitole je popsán vlastní mechanismus generování povrchové formy věty z jejího významového (tektogramatického) zápisu. Třetí kapitola popisuje stručně implementaci a evaluaci celého systému. Po obligátním závěru následují přílohy s ukázkami vět a speciálně i s ukázkami generovaných číslovek.

Autor použil řadu již existujících nástrojů, z nichž nejdůležitější je nástroj pro zpracování stromových struktur ve formátu PDT (tred), v němž práci implementoval. Použité nástroje autor v práci uvedl; jinak vše, co je popsáno v textu práce (kap. 2 až 4), je výsledkem vlastní práce autora.

Hlavním problémem, který autor řešil, bylo rozdělení úlohy na jednotlivé kroky (modularizace). Celá úloha je natolik komplexní, že tento krok byl naprosto nezbytný. Analýzou problému a experimentováním se specifikací generování posléze autor dospěl k několika bodům (a v jejich rámci k dalším podbodům), které je možno řešit relativně samostatně. Tuto analýzu a zejména z ní vzniklé rozdělení považují za velmi cenný výsledek práce.

Autor generování na základě výše uvedené analýzy i plně implementoval a vyhodnotil podle světově uznávané metriky (BLEU). Dosáhl přitom výrazného zvýšení kvality generování oproti jednoduchému základnímu algoritmu (tzv. baseline), a i subjektivně lze výsledné vygenerované věty považovat za velmi dobré. Je ovšem jasné, že ve vlastním strojovém překladu (při neúplném nebo částečně chybném vstupu do procesu generování) budou výsledky horší, to však nebylo cílem diplomové práce.

Ve vlastní práci je v kapitole „Implementace“ pouze popis instalace; popis implementace je až ve vlastních zdrojových programech (na přiloženém CD). Bylo by vhodné popsat způsob implementace alespoň na úrovni modulů z kap. 2. Závěr práce je rovněž velmi stručný (byť vtipný); tato práce by zasloužila delší pojednání o tom, jak je z pohledu autora ještě možno (a za jakého předpokládaného úsilí) v práci dále pokračovat, které jevy bude možno zpracovat snadno a které vyžadují více úsilí nebo dokonce změnu v definici a/nebo naplnění výchozí datové struktury (tj. tektogramatické roviny).

Práce je psána česky, po formální stránce je v pořádku, včetně odkazů na použité prameny a uvedení převzatého software. Výsledky práce budou v nejbližší době využity v projektu strojového překladu z angličtiny do češtiny jak v rámci národních, tak v rámci mezinárodních projektů ÚFAL.

Závěr: Diplomová práce Jana Ptáčka splňuje všechna kritéria kladená na diplomovou práci a jako takovou ji doporučuji uznat.

Praha, 1.2.2006

Výpracoval: Jan Hajič, ÚFAL MFF UK