

Posudek vedoucího diplomové práce
Josef Čech: Spojování segmentů v českých souvětích

Diplomová práce Josefa Čecha se zabývá pojmem segmentu v českých souvětích a studuje možné metody jejich spojování. Toto téma navazuje na předchozí práce a články a snaží se posunout zpracování segmentů od jejich detekce k analýze a spojování do klauzí. Hlavním úkolem diplomové práce bylo vyvinout alespoň dvě odlišné metody spojování a vzájemně je porovnat.

Diplomová práce je napsána česky, je přehledně uspořádaná a skládá se z deseti kapitol, z nichž první slouží jako úvod a poslední jako závěr. V úvodních kapitolách autor postupně uvádí čtenáře do problematiky prostřednictvím úvodního nastínění problému, přes seznámení se základními pojmy, až po obsáhlé vysvětlení procesu segmentace a souvisejících otázek. Teprve šestou kapitolou začíná popis autorova vlastního výzkumu. Tato kapitola přehledným způsobem vysvětluje problémy při získávání segmentů a jejich značkování.

V sedmé kapitole autor analyzuje vlastní téma práce – metody spojování segmentů do klauzí. V souladu se zadáním studuje dva základní přístupy, pravidlový a statistický. U pravidlového přístupu autor obsáhle vysvětluje důvody, které jej vedly k rozhodnutí nevytvářet obecný formalismus psaní pravidel pro spojování, ale zařadit tato pravidla přímo do kódu programu provádějícího spojování segmentů. S uváděnými důvody je možné souhlasit. U přístupu založeného na datech autor uvádí důvody, které jej vedly k použití skrytých markovovských modelů.

Osmá kapitola popisuje implementaci programu na spojování klauzí, jenž má část společnou pro oba přístupy, na kterou navazují dvě varianty podle zvoleného přístupu. Toto rozdělení je logické. První experimenty ukázaly velmi nízkou úspěšnost pravděpodobnostního přístupu, proto se autor v této kapitole zabývá i jeho rozšířeními.

Devátá kapitola obsírně rozebírá dosažené výsledky. Bohužel se autorovi nepodařilo výrazněji zlepšit slabé výsledky pravděpodobnostního přístupu, který i po řadě vylepšení zaostává za přístupem pravidlovým. Výsledky jsou dokumentovány velkým množstvím tabulek, které jsou v textu této kapitoly podrobně popsány a rozebrány. V závěrečné desáté kapitole autor shrnuje dosažené výsledky.

Na práci je nutné ocenit především rozsah provedených experimentů a fakt, že se autor musel vypořádat s celou řadou potíží, které vyvstaly při analýze problému a se kterými zadání diplomové práce nepočítalo. Zvolená řešení jsou adekvátní a logicky zdůvodněná. Určitým nedostatkem práce je to, že se autorovi nepodařilo dosáhnout srovnatelně úspěšných výsledků oběma metodami. Bohužel ale nelze zároveň odmítnout pravděpodobnostní přístup jako méně vhodný, protože z práce není jasné, zda nižší dosažená úspěšnost není způsobena nevhodnou aplikací statistických metod. Pro potvrzení nebo vyvrácení této hypotézy by pomohly další experimenty. Přesto ale mohu konstatovat, že po obsahové stránce práce splňuje zadání.

Po formální i typografické stránce práce také splňuje veškeré požadavky kladené na diplomovou práci, a proto ji doporučuji k obhajobě.

Praze dne 19.5.2014

Vladislav Kuboň
ÚFAL MFF UK