

Strukturovaný posudek bakalářské práce

Milan Nikl

Odhad morfologických značek pro neznámá slova

1. Informace k zadání

Zadání vzniklo z potřeby LIKS (Laboratoř inteligentních komunikačních systémů) na vylepšení úspěšnosti stávajícího morfologického značkovače. Bakalářská práce se zabývá morfo-syntaktickou analýzou neznámých (předem neviděných slov) s nimiž je obecně těžké pracovat. Na úvod je třeba říci, že se jedná o nadstandardně náročné zadání, neboť vyžadovalo od autora pochopení poměrně náročných matematických modelů.

2. Aktivita během řešení, konzultace, komunikace – 0 bodů (max. 15 bodů)

Bakalář přistupoval k řešení velmi nespolehlivě a neaktivně. Opakovaně nedodržoval termíny a dílčí části práce odevzdával pozdě. Práce byla dokončena s výrazným zpožděním. Toto představuje pro vedoucího práce značné zklamání. Vzhledem k tomu, že výsledný produkt nakonec splňuje požadavky, považuji výše uvedené problémy za poněkud zbytečné a doufejme, že si bakalář tyto výtky vezme k srdci.

3. Splnění požadavků zadání – 25 bodů (max. 25 bodů)

Práce splňuje zadání a navíc byla rozšířena o experimenty na slovenském jazyce.

4. Hodnocení formální stránky předložené práce – 25 bodů (max. 25 bodů)

Dokument je vysázen v systému LaTeX a je díky tomu vkusně upraven. Autor dodržuje typografické konvence, jeho práce je dobře čitelná a srozumitelná. Vyjadřování v odborném českém jazyce je také na úrovni. První třetina práce se zabývá popisem problému z lingvistického hlediska. Důkladně popisuje různé jazykové záludnosti, se kterými je nutné počítat při přidělování morfologický značek. Další třetina poskytuje matematický aparát pro výpočet pravděpodobností přidělovaných značek. Zbytek dokumentu pak popisuje způsob implementace a podrobné výsledky experimentů na českém a slovenském jazyce.

5. Hodnocení realizačního výstupu – 30 bodů (max. 35 bodů)

Bakaláři se podařilo zvýšit úspěšnost stávajícího systému o přibližně 2,5% na českém i slovenském jazyce což považuji za dostatečný přínos. Výsledný program je napsán v jazyce Java a je plně funkční, avšak mám několik výhrad ke stylu programování. Autor míchá ve zdrojových kódech češtinu i angličtinu, což působí poněkud rušivě. Také nepovažuji za šťastné, že skoro všechny metody jsou *static*, což poněkud odporuje objektovému programování v jazyce Java.

6. Otázky k obhajobě

K vlastní práci mám dvě otázky:

- V práci prezentujete, že se Vám podařilo zrychlit značkování o přibližně 80%. Čím si to vysvětlujete?
- Dal by se Váš způsob odhadování značek neviděných slov použít i na další jazyky?

7. Závěrečné shrnutí – celkem dosaženo 80 bodů (max. 100 bodů)

Práci doporučuji k obhajobě.

V Plzni dne 20.8.2012

Ing. Tomáš Bryhcín
KIV – FAV - ZČU

