

Posudek oponenta diplomové práce

Diplomová práce **Jana Bušty** se zabývá extrakcí slovesných idiomů z korpusu.

Text práce je místy příliš stručný až heslovitý. V textu jsou použity anglické termíny (např. od str. 10 *token-based/type-based extraction*), které přitom mají přirozené české ekvivalenty. Někdy text obsahuje poněkud neobratné formulace, které mohou poskytovat zavádějící informace, např.:

- (str. 34) *algoritmus odhalil 60 až 70 % idiomů*, přitom uvedená procenta udávají počet idiomů z navržených frází,
- (str. 29) *každé sedmé slovo je součástí nějaké slovesné idiomatické fráze*, to by znamenalo, že každá věta obsahuje idiomatickou frázi.

Domnívám se, že práci chybí nějaká přesnější definice idiomů. Bez ní je práce anotátorů a tím i celé vyhodnocení poněkud nepřesné. Osobně se mi nezdá jako idiom např. *polykat prášek* (str. 24), či *sklonit hlavu* (str. 27), naopak jako idiom bych označil např. *mít opici* (str. 20).

K práci mám následující dotazy:

- Na str. 14 je uvedeno, že se pracuje se zhruba 45 tisíci spojeními, na str. 19 je potom uveden počet potřebných dotazů 1,2 milionu. Jak byl tento počet určen?
- Jak byl použit thesaurus, zejména jak byla použita hodnota skóre pro jednotlivá slova?

Zadání práce bylo splněno. Práci hodnotím známkou *C–D* podle výsledků obhajoby.

Ve Vranově , 1. února 2012

Pavel Rychlý