

## Posudek oponenta bakalářské práce

**Student:** Haljuk Petr, Bc.  
**Téma:** Shlukování slov podle významu (id 17717)  
**Oponent:** Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno  
Zadání bylo splněno, student vytvořil systém, který dokáže k zadanému slovu automaticky najít slova sémanticky příbuzná.
3. **Rozsah technické zprávy** je v obvyklém rozmezí  
Rozsah technické zprávy je v obvyklém rozmezí, všechny její části jsou informačně bohaté a pro práci nezbytné, snad až na časté ukázky zdrojových kódů.
4. **Prezentační úroveň předložené práce** 70 b. (C)  
Logická struktura práce je dobrá, rozsahy a návaznosti jednotlivých kapitol také. Práci ale trpí občasnými nedostatky. Některé pojmy zůstávají v práci nevysvětleny. Některá tvrzení jsou nepřesná nebo sporná. Práce se také omezuje na vybrané algoritmy, zatímco jiné, běžně používané, nejsou v práci popsány (LSA, LDA atd.).
5. **Formální úprava technické zprávy** 70 b. (C)  
Po formální stránce lze práci vytknout několik překlepů, pravopisných chyb, občasnou chybějící diakritiku v obrázcích, občasně nevhodně formulované věty, případně nevhledné a nečíslované ukázky zdrojových kódů.
6. **Práce s literaturou** 70 b. (C)  
Výběr studijních pramenů je vhodný. Nicméně některé webové odkazy na použité technologie se hodí spíše do poznámky pod čarou než do seznamu použité literatury. Vytkl bych také použití Wikipedie, která se v seznamu literatury nachází hned čtyřikrát. Student se na tyto prameny v práci vhodně odkazuje.
7. **Realizační výstup** 80 b. (B)  
Realizační výstup je na dobré úrovni. Dle všeho by se implementovaný systém umístil někde kolem 20. místa ze 38 v mezinárodní soutěži SemEval 2014.
8. **Využitelnost výsledků**  
Realizační výstup práce je možné použít k hledání podobných slov k danému slovu, případně k účasti na budoucích soutěžích SemEval.
9. **Otázky k obhajobě**
  - Proč jste zvolil velikost vektoru 300 a velikost kontextového okna 5 pro metodu CBOW, resp. 10 pro Skip-gram? Zkoušel jste s těmito hodnotami nějak experimentovat? Pokud by tyto hodnoty byly vyšší, dosažené výsledky by byly pravděpodobně lepší.
  - Jaké rychlosti jste dosahoval při trénování modelu? Spouštěl jste trénování paralelně? Jaká byla rychlost výpočtu podobných slov?
10. **Souhrnné hodnocení** 75 b. dobře (C)  
Student implementoval systém pro výpočet podobnosti slov. Systém dosahoval lehce nadprůměrných výsledků v porovnání s výsledky na soutěži SemEval 2014. Technická zpráva měla několik výše zmíněných nedostatků. Z těchto důvodů navrhuji hodnocení stupněm C.

V Brně dne: 3. června 2015

.....  
podpis