



## OPONENTNÍ POSUDEK ZÁVĚREČNÉ KVALIFIKAČNÍ PRÁCE

**Autor závěrečné práce:** Bc. Radek Šafařík

**Název práce:** Metody měření podobnosti jazyků

**Oponent práce** Mgr. Jan Macošek

**Pracoviště opONENTA** Newton Media, a. s.

- |                                                                                                     |                    |
|-----------------------------------------------------------------------------------------------------|--------------------|
| A. Kvalita abstraktu, klíčová slova odpovídají náplni práce . . . . .                               | Výborně (1)        |
| B. Rozsah a zpracování rešerše . . . . .                                                            | Výborně mínus (1-) |
| C. Řešení práce po teoretické stránce . . . . .                                                     | Výborně (1)        |
| D. Vhodnost, přiměřenost použité metodiky . . . . .                                                 | Výborně (1)        |
| E. Úroveň zpracování výsledků a diskuse . . . . .                                                   | Výborně (1)        |
| F. Vlastní přínos k řešené problematice . . . . .                                                   | Výborně (1)        |
| G. Formulace závěru práce . . . . .                                                                 | Výborně (1)        |
| H. Splnění zadání (cílů) práce . . . . .                                                            | Splněno            |
| I. Skladba, správnost a úplnost citací literárních údajů . . . . .                                  | Výborně (1)        |
| J. Typografická a jazyková úroveň (vč. pravopisu) . . . . .                                         | Velmi dobře (2)    |
| K. Formální náležitosti práce . . . . .<br>(struktura textu, řazení kapitol, přehlednost ilustrací) | Výborně (1)        |

Komentáře či připomínky:

...pokračuje na straně 2



**Celkové zhodnocení:**

Diplomová práce Bc. Radka Šafaříka zpracovává téma měření podobnosti jazyků, nikoli ovšem z pohledu čistě lingvistického, ale za použití statistických metod a měření. Jako taková poskytuje cenný pohled na evropské (a zejména slovanské) jazyky, výsledky měření pak lze využít nejen k lingvistickému výzkumu, ale i k výraznému usnadnění práce na systémech pro zpracování řeči.

Diplomant musel zpracovat a implementovat nejen metody samotného měření, ale také převodů ortografické podoby jazyka do fonetické. To vše následně aplikoval na 6 paralelních korpusů (čítajících až 47 jazyků) a výsledky pak zejména pro češtinu velmi podrobně a čitelně sumarizoval.

Jediná výtka směřuje ke stránce typografické a jazykové, konkrétně k občasné chybě fontu v ukázkách z korpusů a několika překlepům. Tyto drobné vady ale nemůžou zastínit vysokou obsahovou kvalitu práce.

**Otázky k obhajobě:**

1. Uvažoval jste o omezení vlivu flexe jednotlivých jazyků, např. použitím lemmatizátorů? Myslíte, že by jejich použití výrazněji ovlivnilo naměřené podobnosti?
2. Který jazyk, dosud nezpracovaný v Laboratoři počítačového zpracování řeči na TUL, byste na základě svých výsledků doporučil k budoucímu zpracování (s využitím českých modelů)?

**Celková klasifikace:**

Práce splňuje požadavky na udělení akademického titulu, a proto ji doporučuji k obhajobě

Navrhuji tuto práci klasifikovat stupněm **Výborně (1)**

V Praze

dne 2. 6. 2014

Podpisem současně potvrzuji, že nejsem v žádném osobním vztahu k autorovi práce

. . .  . . . . .

podpis oponenta