

# Posudek vedoucího diplomové práce

Autor práce: **Bc. Petr Kopal**

Název práce: **Propojení témat zpravodajských článků mezi jazyky**

Cílem práce bylo nastudovat metody pro výpočet podobnosti mezi texty v různých jazycích. Dále pak navrhnout metodu, která dostane jednojazyčné shluky článků ze systému pro monitoring médií a propojí související napříč jazyky.

Nejprve diplomant popisuje základy NLP a vybrané metody distribuční sémantiky. Další část již pojednává o výpočtu kroslinguální podobnosti textů a jejich shlukování. Na následujících cca 20 stranách lze nalézt popis implementace software, který byl v rámci diplomové práce vytvořen. V závěru jsou diskutovány experimenty a dosažené výsledky.

Baseline metoda, použití deskriptorů z thesauru Eurovoc a pojmenovaných entit, byla v rámci diplomové práce překonána přidáním metody CL-LSA. V rámci experimentů byly zkoumány dvě varianty: jednojazyčné shlukování + linkování a krosjazyčné shlukování. Lepších výsledků dosáhla první varianta. Implementace LSA zvládla zpracovat korpus o velikosti maximálně jednotek desetitisíc dokumentů. Dimenze LSA byla redukována na 500, což by zasloužilo další testy nebo zdůvodnění.

Diplomant na tématu pracoval dva roky, ačkoliv většina práce byla odvedena v závěrečném finiši. Nutno říci, že se jednalo o složité výzkumné téma, do kterého nebylo jednoduché proniknout bez předchozích znalostí zpracování přirozeného jazyka. Navržené řešení má své rezervy a dospět k tomu, aby bylo publikovatelné ve výzkumné komunitě, by znamenalo nemalá rozšíření a další experimenty. Nicméně hlavně díky výraznému zvýšení úsilí v posledních třech měsících se podařilo nakonec odevzdat obstojné dílo.

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 20. 8. 2019

Doc. Ing. Josef Steinberger, Ph.D.