

## Posudek vedoucího/opponenta\* diplomové práce

Jméno a příjmení autora posudku: RNDr. Ondřej Bojar, Ph.D.

Jméno a příjmení autora práce: Duc Tam Hoang

Název práce: Pivoting Machine Translation for Vietnamese

---

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

The thesis submitted by Tam Hoang aims at developing statistical machine translation systems between Czech and Vietnamese. While both Czech and Vietnamese alone have reasonably developed NLP data and resources, as a pair they fall into the low-resource category. The thesis thus improves the translation quality by bringing in English as the pivot language.

The thesis explores several methods of this so-called 'pivoting'. The improvement over the baseline direct translation is achieved not by the individual methods but only when they are combined with the baseline. An inherent and important part of the thesis was data collection and cleanup, which proved very effective for translation between Vietnamese and Czech.

The thesis is well organized and provides all the details necessary to replicate the experiments. I find very useful the graphical summaries of the various setups, e.g. Fig. 6.1 or 6.4, which give a clear overview of otherwise rather complex set of the building blocks of the experiments. All the results are well commented and discussed.

The main shortcoming of the thesis is the language. As the supervisor, I spent a lot of time on improving Tam's grammar and style. This helped but the text still suffers at many places. It should be however stressed that the errors are not an obstacle to understanding the message.

What I value most is the range and number of conducted experiments. While the methods were previously suggested in the literature (which the thesis carefully cites) and despite limited to one language pair (as required), the description and results can serve as a generic cookbook for pivoting methods. All the methods were tested in a comparable setting, wideing the scope of previous works in this field. Tam has also released his implementation for one of the methods, phrase-table triangulation, as an open-source tool accompanied with an MT Marathon/PBML paper (in print). In contrast to the need for supervision with English grammar, Tam needed very little guidance in the scientific experimenting. From the research point of view, Tam is clearly capable of designing and conducting experiments on his own. Therefore I can confidently recommend the thesis to be accepted as a M.Sc thesis at Charles University in Prague.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input type="checkbox"/>
---	------------------------------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

--

V Praze dne: 28. srpna 2015

Podpis:\*\*

*\* nehodící se škrtněte (vymažte)*

*\*\* do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*