

## Posudek disertační práce

Posudek vedoucího

Řešitel: Mgr. Aleš Tamchyna  
Název práce: Lexical and Morphological Choices in Machine Translation  
Vedoucí: RNDr. Ondřej Bojar, Ph.D.

Aleš Tamchyna se ve své disertační práci zabývá frázovým statistickým strojovým překladem s cílem vylepšit kvalitu překladu ve dvou komplementární otázkách: výběru překladových ekvivalentů (lexikální volba) a výběru správného tvaru slov (morfologická volba).

Výchozí model, frázový statistický překlad byl donedávna obecně nejpoužívanější; pro svoji obecnost a prakticky nulové požadavky na lingvistické předzpracování textu byl široce použitelný a napříč jazykovými páry a doménami dosahoval v průměru nejlepších známých výsledků. Aleš ve své práci staví na sofistikované sestavě, tzv. systému Chiméra, který frázový překlad kombinuje s dvěma lingvisticky založenými komponentami, a k němuž Aleš též přispěl přípravou dat a budováním součástí modelu.

Text práce je přehledně organizován do devíti kapitol, po úvodní motivaci je popsán frázový překlad se všemi běžně užívanými technikami (log-lineární model, optimalizace vah, prohledávání stavového prostoru, faktory) a to do hluboké úrovně technického detailu, jak je nezbytné pro vysvětlení vlastního příspěvku. Ve třetí kapitole je stejná pečlivost věnována systému Chiméra, kde se Aleš významně podílel na analýze důvodů jeho úspěšnosti, a kapitola slouží též jako velmi podrobné představení hlavních omezení frázového modelu, která Aleš svou prací částečně odstraňuje. Čtvrtá kapitola shrnuje základy strojového učení nutné pro pochopení Alešova vlastního příspěvku. Alešův stěžejní příspěvek předkládané práce, diskriminativní frázový překladový model (DPTM), je pak popsán v kapitole páté. Šestá kapitola je věnována integraci DPTM do překladového systému, kde se jako zásadní ukazují Alešovy optimalizace výpočtu, aby byl model v praxi použitelný a mj. dovoloval zapojit kontext cílové věty. Sedmá kapitola přináší empirické vyhodnocení mnoha konkrétních konfigurací modelu pro překlad z angličtiny do češtiny a též do němčiny a rumunštiny. Ve všech případech DPTM významně zlepšuje automatická skóre, překlad do češtiny byl vyhodnocen též ručně a DPTM pomáhá, pokud se opírá o zdrojový i cílový kontext. Práce je zakončena přehledem souvisejících prací (kapitola 8) a závěrem.

Je třeba poznamenat, že v průběhu posledního roku řešení práce zaznamenal obor strojového překladu zásadní změnu paradigmatu. V průběhu jediného roku se podařilo kolegům z univerzity v Montrealu a Edinburku překonat hned několik velkých překážek a zprovoznit tzv. neuronový strojový překlad na velkých datech. Neuronový překlad s kontextovou informací pracuje zcela odlišně a v zásadě netrpí problémy, k jejichž řešení Alešův přístup přispěl ve frázovém překladu. V tak pokročilé fázi doktorského studia ovšem nebylo realistické zcela změnit téma. Aleš se, podobně jako celá komunita, rychle adaptoval (což dokládá i svým příspěvkem zlepšujícím neuronový překlad zaslaným na konferenci EMNLP 2017), ale začít pracovat na disertační práci z nuly nedávalo dobrý smysl.

Jakkoli se v současné době studuje již prakticky jen neuronový překlad, frázový překlad nezřídka empiricky vychází lépe, zejména v oblastech s dostatkem trénovacích dat v relativně pevné doméně. Taková situace je běžná pro průmyslové překlady, a Alešův příspěvek je tedy stále velmi relevantní, byť již ne pro obecný překlad bez omezení domény.

Práce je psána vynikající angličtinou a po formální stránce je prakticky bez chyb. Použité práce jsou vždy řádně citovány a jejich seznam je více než dostatečný, prokazuje autorův široký přehled o starších i nejnovějších poznatcích.

Závěrem jednoznačně konstatuji, že předložená práce Mgr. Aleše Tamchyny splňuje nároky kladené na disertační práce a proto ji **doporučuji k přijetí**.

Praha, 26. květen, 2017.

RNDr. Ondřej Bojar, Ph.D.