

Review of Master's Thesis

Student: Sýkora Tomáš, Bc.
Title: The Best Possible Speech Recognizer on Your Own Data (id 18056)
Reviewer: Veselý Karel, Ing., Ph.D., DCGM FIT BUT

- 1. Assignment complexity** **considerably demanding assignment**
Zadání je náročné. Složitost zadání spočívá zejména : 1) v náročné "studijní fázi" nutné pro proniknutí do oblasti rozpoznávání řeči, 2) potřebě seznámení se s toolkitem Kaldi a dalšími použitými toolkity, 3) v pracné přípravě řečových dat, jejich přepisů, textových dat a výslovnostního slovníku, 4) v experimentální práci, zejména v jazykovém modelování, ale i stavbě rozpoznávače celkově.
- 2. Completeness of assignment requirements** **assignment fulfilled with enhancements**
Rozšíření nad rámec zadání jsou: experimenty s vkládáním ticha na začátky a konce řečových segmentů, trénování pravděpodobnosti ticha mezi slovy, experimenty s jazykovým modelováním (interpolace, prořezávání, reskórování pomocí RNNLM).
- 3. Length of technical report** **in usual extent**
Rozsah práce je standardní a je dobře, že jde "k věci" a neobsahuje nadbytečné části.
- 4. Presentation level of technical report** **80 p. (B)**
Jedná se o čtivý text, který vhodně vystihuje problematiku zadání, které je zaměřeno na stavbu rozpoznávače řeči z vlastních dat. Členění do kapitol je logické a vhodně zvolené. Na druhou stranu mi v práci chybělo uspokojivé vysvětlení jak "uvnitř" funguje dekodér. A experimentální část je místy obtížněji srozumitelná. Nicméně jednotlivostí, které jsou v práci popsány, je velké množství. A tak rozumím tomu, že ne vše bylo možné popsat detailně.
- 5. Formal aspects of technical report** **95 p. (A)**
Typografická a jazyková úroveň práce je na vysoké úrovni.
- 6. Literature usage** **100 p. (A)**
V práci je citováno 43 vhodně zvolených pramenů. Student využil mnoho relevantních zdrojů.
- 7. Implementation results** **80 p. (B)**
Autor přiložil skripty, které použil pro přípravu dat, trénování systému i pro demonstraci nasazení v on-line režimu. Skripty používají ale neobsahují toolkity Kaldi (licence Apache 2.0) a KenLM (licence LGPL 2.1).
- 8. Utilizability of results**
Toto je velmi silná stránka této práce. Realizační výstup je rozpoznávač řeči, který je/bude nasazený firmou Parrot pro přepis justičních dat. Díky této diplomové práci se firmě Parrot počet chyb přepisu snížil téměř na polovinu, z 37.4% na 20.2%.
- 9. Questions for defence**
 - Q1: Vysvětlete princip fungování log-lineární interpolace jazykových modelů. Co se děje se skóre častých a málo častých slov, když má jazykový model negativní exponent? (např. v tabulce 5.2)
 - Q2: Která oblast praktické části byla nejpracnější a která další byla náročná?
- 10. Total assessment** **90 p. excellent (A)**
Práce je celkově zdařilá. Jedná se o pracné a dobře provedené inženýrské dílo s možností okamžitého využití. Práce dále obsahuje experimentální rozšíření, která jdou nad rámec zadání.

In Brno 29 June 2020

Veselý Karel, Ing., Ph.D.
reviewer