



Experimentálny rozpoznávač spontánnej reči

Diplomový seminár 1

Školiteľ:

RNDr. Marek Nagy, PhD.

Andrej Kapusta

14. 05. 2021

Zadanie práce

- Vytvorenie rozpoznávaču spontálnej reči
 - Nahrávky z interview, prezentácií a.i.
 - Vytvorenie prepisu
- Spolupráca s Transparency International Slovensko (TISK)



Priebeh práce v LS

- Hĺbkové oboznámenie sa s témom práce a analýza domény
- Analýza open-source ASR toolkit-ov
 - HTK toolkit
 - CMUSphinx
 - Kaldi
- Získanie a spracovanie nahrávok od TISK
- Vykonanie prvého experimentu v nástroji Kaldi

Kaldi toolkit

- Open-source nástroj
- Automatické rozpoznávanie hlasu
- Podporuje:
 - HMM (Hidden Markov Models)
 - DNN (Deep Neural Networks)
- Aktívny vývoj
- Licencia Apache 2.0
- Obsahuje iné projekty a ukázky trénovania a testovania



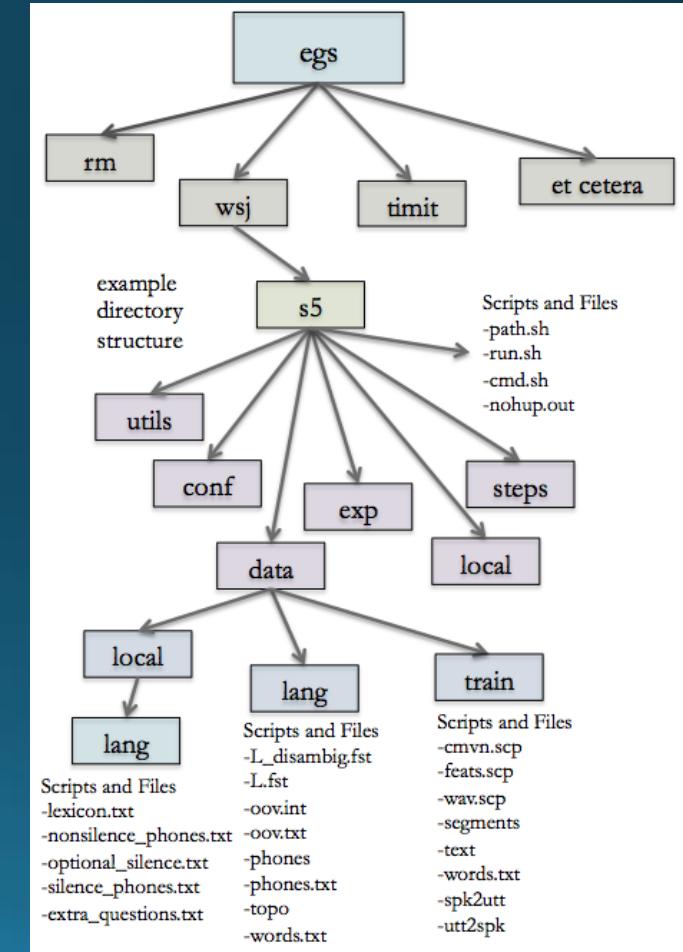
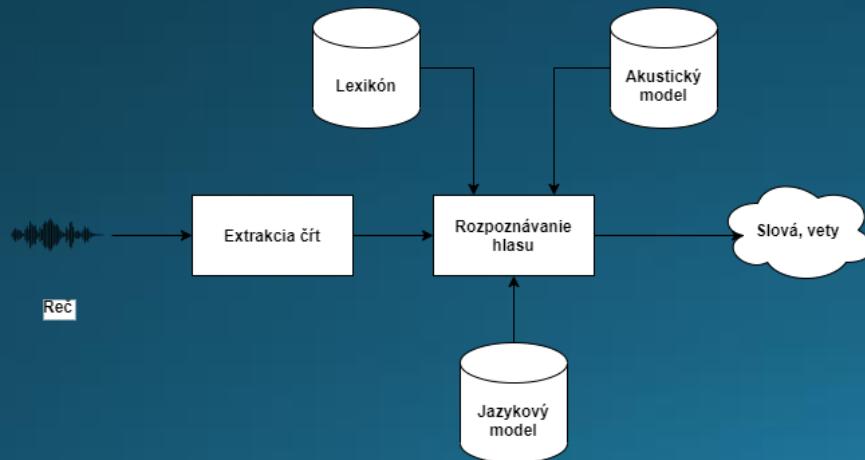
Experiment: Príprava nahrávok a prepisov

- Požiadavky toolkitu na typ nahrávok:
 - Typ audio súboru: .wav
 - Vzorkovacia frekvencia: 16Khz
 - Počet kanálov: 1 (mono)
- Spracovanie nahrávok a prepisov od TISK
- Dva scenáre experimentu
 - Počiatočný dataset (1, 2)
 - Rozšírený dataset (1, 2, 3)

Číslo nahrávky	Pred spracovaním	Po spracovaní
1	00:16:31	00:01:47
2	00:12:48	00:01:29
3	00:12:51	00:01:54

Experiment - vytvorenie projektu

- Využitie postupu od autorov toolkitu Kaldi [6]
- Vytvorenie súborovej štruktúry projektu
- Vytvorenie súborov pre akustický model
- Vytvorenie súborov pre jazykový model
- Príprava trénovacích skriptov



Experiment – akustický model

- spk2gender
- utt2spk

```
moderator1 f  
moderator2 f  
recnik1 m  
recnik2 f
```

```
moderator1-03 moderator1  
moderator1-04 moderator1  
moderator1-05 moderator1  
moderator2-01 moderator2  
moderator2-02 moderator2
```

- wav.scp

```
moderator1-03 /home/andrej/Desktop/Kaldi/kaldi/egs/experiment/data_audio/train/moderator1/exp-3-f-moderator1.wav  
moderator1-04 /home/andrej/Desktop/Kaldi/kaldi/egs/experiment/data_audio/train/moderator1/exp-4-f-moderator1.wav  
moderator1-05 /home/andrej/Desktop/Kaldi/kaldi/egs/experiment/data_audio/train/moderator1/exp-5-f-moderator1.wav  
moderator2-01 /home/andrej/Desktop/Kaldi/kaldi/egs/experiment/data_audio/train/moderator2/exp-1-f-moderator2.wav  
moderator2-02 /home/andrej/Desktop/Kaldi/kaldi/egs/experiment/data_audio/train/moderator2/exp-2-f-moderator2.wav
```

- text

```
moderator1-03 keďže si pracoval s tými dátami a to bol v podstatne taký veľký súbor tých dát našiel si v nich niečo  
moderator1-04 ako dlho trvala celá tvoja stáž  
moderator1-05 sú ešte nejaké iné úlohy ktoré osobne si mal preferenciu  
moderator2-01 čo ta motivovalo aby si sa stala dobrovoľníčkou pre transparency  
moderator2-02 a prečo si si vybrała práve transparency ako organizáciu
```

Experiment – jazykový model

- lexicon.txt

```
druhou d r u h ɔu
dva d v a
dvoch d v ɔ x
emu ε m u
ešte ε ſ c ε
falošné f a l ɔ ſ n ε:
funguje f u ŋ g u j ε
hľadani h λ a d a n i:
hlásila h l a: s I l a
```

- silence_phones.txt

```
sil
spn
```

- nonsilence_phones.txt

```
a
a:
æ
b
c
d
dʒ
dʒ
ɛ
ɛ:
```

- corpus.txt

Experiment – výsledky

- Validácia na trénovacích dátach

Model	Počiatočný dataset WER [%]	Rozšírený dataset WER [%]
Mono	28.27	14.67
TR1	17.56	10.46

- Validácia na testovacích dátach

Model	Počiatočný dataset WER [%]	Rozšírený dataset WER [%]
Mono	87.22	29.55
TR1	87.97	32.95

Postup ďalšej práce



[5]

- Akvizícia ďalších zvukových dát a prepisov
- Preskúmanie možností využitia DNN (deep neural networks)
- Vykonanie druhého experimentu v nástroji Kaldi (s použitím DNN)
- Dopísanie teoretickej časti

Ďakujem

Priestor na otázky

Zdroje

- [1] https://transparency.sk/wp-content/uploads/2020/11/tis_full-1.svg
- [2] <https://www.voicereport.ai/wp-content/uploads/2019/08/speech-to-text-software.png>
- [3] https://kaldi-asr.org/kaldi_text_and_logo.png
- [4] <https://www.eleanorchodroff.com/tutorial/kaldi/familiarization.html>
- [5] https://www.iconfinder.com/icons/6585312/checklist_document_list_menu_icon
- [6] https://kaldi-asr.org/doc/kaldi_for_dummies.html