

SEASR-CZE
ROZPOZNÁVAČ ŘEČI PRO ÚČELY
VYHLEDÁVÁNÍ V ČESKY MLUVENÉ
ČÁSTI KORPUSU
uživatelská a instalační příručka



MINISTERSTVO
KULTURY

Obsah

1 Úvod	2
2 Použití	3
3 Instalace	3

1 Úvod

SEASR-CZE slouží ke zpracování česky mluvených nahrávek. Vytváří z nich databázi, se kterou MCLASS (WFBAS) dále pracuje.

Standardní systém rozpoznávání řeči sestává z akustického modelu, modulu pro parametrizaci řeči a jazykového modelu. Akustické modely v našem systému jsou založeny na architektuře skrytých Markovových modelů (HMM), která představuje "lege artis" přístup v současném rozpoznávání mluvené řeči. Jsou použity standardní třístavové akustické modely s Gaussovskými směsmi. Model bere v úvahu trifónové závislosti včetně mezislovních. Řeč je parametrizována pomocí 15PLP koeficientů a jejich delta a delta-delta koeficientů (tj. vektor příznaků má dimenzi 45). Příznaky jsou extrahovány 100krát za vteřinu a je aplikována kepstrální normalizace na úrovni řečníka.

Systém obsahuje též zobecněný model ticha a při jeho tvorbě byly použity špičkové metody pro adaptivní a diskriminativní trénování.

Pokud se týká jazykového modelu, systém obsahuje dva základní modely – oba trigramové. První z nich je natrénován pouze z pomocí přepisů z „mateřské“ domény, zatímco druhý využívá jako trénovací data texty z rozsáhlého Českého národního korpusu.

Vlastní řečový dekodér může být použit v dvouprůchodovém režimu (stále ale zajišťujícím zpracování v reálném čase) – v prvním průchodu se většinou používá obecný jazykový model a ve druhém jsou přeskórovány získané mřížky modelem sofistikovanějším. Dekodér generuje slovní a fonémové mřížky pro následné efektivní vyhledávání.

Bližší informace o zpracování lze nalézt ve článku¹.

¹*System for fast lexical and phonetic spoken term detection in a Czech cultural heritage archive.* [dokument ve formátu PDF] dostupný z: <http://www.asmp.eurasipjournals.com/content/2011/1/10>

2 Použití

Pomocí software SEASR-CZE se vytváří databáze nahrávek. Vzhledem k tomu, že vytvoření takové databáze je jednorázová akce a během jejího používání již není potřeba software spouštět, je vhodné používat software jako samostatný program.

Předkládaný vstupní soubor (v příkladu `example.wav`) musí mít jeden kanál, vzorkovací frekvenci 16kHz a rozlišení 16 bitů.

Příklad použití:

```
Recognition.exe -set AMALACH_CZ.SET -file example.wav -result example.mlf
```

Software byl navržen speciálně pro zpracování české části archivu AMALACH a jeho použití pro jiná data bez dalších úprav se nepředpokládá.

3 Instalace

Software se pouze kopíruje na požadované umístění.