

The logo for 'sages' features the word in a lowercase, sans-serif font. The letter 's' is grey, 'a' is blue, 'g' is grey, 'e' is blue, and 's' is grey. A small blue square is positioned below the 'e'.

Task 5: Automatic speech recognition

PolEval 2019 competition

Danijel Koržinek danijel@pja.edu.pl

May 31st 2019

Task

- Convert a set of audio files into text
- Domain: Polish parliamentary speech
- Audio: WAV, 16 kHz, 16-bit audio from
- Text: UTF-8, lowercase, no punctuation, no numbers, no abbreviations

- Open vs fixed competition
- Provided data:
 - Clarin-PL speech corpus
 - PELCRA parliamentary corpus
 - A collection of 97 hours of parliamentary speeches published on the ClarinPL website
 - Polish Sejm Corpus for language modeling
- In all cases, data from the Sejm/Senat government websites after January 1st 2019 was not to be used

Evaluation method

- A set of audio files from a random day after January 1st 2019 was picked
- The files were processed and manually corrected
- Evaluation measure - Word Error Rate:

$$WER = \frac{N_{del} + N_{sub} + N_{ins}}{N_{ref}} \quad (1)$$

- Calculation based on the NIST Sclite tool

Participating systems

Team	System	Type
MS	GOLEM	Fixed
ML	ARM-1	Open
2*AWSR	SGMM2	Open
	tri2a	Open
2*PJATK	clarin-pl/studio	Open
	clarin-pl/sejm	Fixed

Final results

System	WER%	CORR%	SUB%	DEL%	INS%
GOLEM	12.8	90.1	6.9	3	2.9
ARM-1	26.4	77	16.5	6.5	3.4
AWSR SGMM2	41.3	65.2	27.1	7.7	6.5
AWSR tri2a	41.8	62.9	26.8	10.3	4.7
clarin-pl/studio	30.9	71.4	16	12.6	2.4
clarin-pl/sejm	11.8	89.7	5.4	5	1.4

Per-file statistics

System	Mean	StdDev	Median
GOLEM	13.3	8.8	11.9
ARM-1	27.2	13.5	24.7
AWSR SGMM2	41.3	18.1	38.8
AWSR tri2a	41.4	16.9	38.5
clarin-pl/studio	30.4	13.6	25.9
clarin-pl/sejm	12	7.9	9.8

Conclusion

- **Winner: GOLEM!**
- All systems, except for ARM-1, based on Kaldi
- All systems, except for clarin-pl, using GMM models
- Fixed systems were the only using in-domain data
- GOLEM and Clarin-PL systems available on Github

Thank you for the participation