

Otwartoźródłowe narzędzia rozpoznawania mowy

Michał Andrzej Woźniak

Istnieje szereg otwartoźródłowych systemów rozpoznawania mowy – CMU Sphinx, Janus, Julius, Kaldi, RWTH ASR czy iAtros. Istnieje też LumenVox, który sam otwartoźródłowy (ani darmowy) nie jest, ale jest dobrze przystosowany do współpracy z wolnym oprogramowaniem.

Projekty te są na różnym stadium rozwoju. CMU Sphinx, Kaldi, Julius i LumenVox wydają się najbardziej zaawansowane i dojrzałe, są też aktywnie rozwijane; iAtros – wygląda na porzucony. CMU Sphinx jest wykorzystywany m.in. przez Mozillę w Web Speech API, i wydaje się najlepszym kandydatem do wykorzystania w Projekcie.

Istnieje też szereg narzędzi ułatwiających generowanie modeli, szkolenie systemu rozpoznawania mowy, i innych dodatkowych działań.

Największym problemem będzie znalezienie korpusu językowego w języku polskim. Można go spróbować wygenerować w oparciu o audiobooki w LibriVox, i narzędzia VoxForge.

Konieczne jest przeprowadzenie testów konkretnych systemów (przede wszystkim: CMU Sphinx, Kaldi, Julius), i ustalenie, które w praktyce najlepiej nadają się do wykorzystania w Projekcie.

Należy jednak nastawić się, że rozpoznawanie mowy będzie zawsze obarczone błędem. Narzędzia rozpoznawania mowy są szkolone na próbkach, wydajność i dokładność spada wraz z:

- wzrostem wielkości słownika (słownik w ramach Projektu musiałby obejmować cały słownik języka polskiego);
- wzrostem liczby mówiących, których mowę należy rozpoznawać (system musiałby rozpoznawać mowę losowych osób).

Jednym z możliwych rozwiązań jest szkolenie systemu do rozpoznawania mowy konkretnej osoby, która podczas konsultacji kwestie mówione przez uczestników powtarzać będzie do dedykowanego mikrofonu w celu ułatwienia rozpoznania mowy.

Systemy rozpoznawania mowy

CMU Sphinx:

- <http://cmusphinx.sourceforge.net/>

Janus:

- [http://en.wikipedia.org/wiki/Janus_Recognition_Toolkit_\(JRTk\)](http://en.wikipedia.org/wiki/Janus_Recognition_Toolkit_(JRTk))

Julius:

- [http://en.wikipedia.org/wiki/Julius_\(software\)](http://en.wikipedia.org/wiki/Julius_(software))
- http://julius.sourceforge.jp/en_index.php?q=en/index.html

LumenVox (zamknięte, płatne! Kompatybilne z WiOO)

- <http://en.wikipedia.org/wiki/LumenVox>

Kaldi:

- <http://kaldi.sourceforge.net/>

RWTH ASR:

- <http://www-i6.informatik.rwth-aachen.de/rwth-asr/>

iAtros:

- <https://www.prhlt.upv.es/software/iatros/doc/speech/>

Narzędzia

HTK Toolkit:

- <http://htk.eng.cam.ac.uk/>

CSU Toolkit:

- <http://www.cslu.ogi.edu/toolkit/>

Simon:

- <https://simon.kde.org/>

Mozilla Web Speech API:

- https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API
- https://wiki.mozilla.org/HTML5_Speech_API

- <https://wiki.mozilla.org/SpeechAPI>
- https://wiki.mozilla.org/SpeechRTC_-_Speech_enabling_the_open_web

Banki danych/próbek/modeli

VoxForge:

- <http://en.wikipedia.org/wiki/VoxForge>
- <http://www.voxforge.org/>

LibriVox:

- <http://en.wikipedia.org/wiki/LibriVox>
- <https://librivox.org/>

TalkBank:

- <http://en.wikipedia.org/wiki/TalkBank>
- <http://talkbank.org/>

Źródła dodatkowe

- <https://fedcsis.org/proceedings/2011/pliks/236.pdf> (CMU Sphinx)
- <http://pe.org.pl/articles/2014/4/42.pdf> (CMU Sphinx)
- http://www.fa2014.agh.edu.pl/fa2014_cd/article/PJ/PJ03_9.pdf (Kaldi)
- <http://spirit.blau.in/simon/tag/polish/> (Simon)