

Development of a cloud platform for automatic speech recognition

Ondřej Klejch

Supervisor: Mgr. Ing. Filip Jurčíček Ph.D.



Goals

- Develop a cloud based speech recognition API.
- Develop an annotation interface.

Motivation

There is a need for a platform that:

- Provides an ASR functionality to non-experts.
- Enables sharing of trained ASR models.
- Allows us to collect more spoken data.
- Scales well with load.

Features

- Built on top of Kaldi.
- Supports batch and online speech recognition.
- Contains an annotation interface for transcription of submitted recordings.
- Scalable, customizable and easily deployable.
- Publicly available at <http://www.cloudasr.com>
- Apache 2.0 license

Impact

- Several companies are interested in CloudASR, because they can ensure data privacy by running the platform on their own machines.
- CloudASR was deployed in Fondazione Bruno Kessler, Trento, Italy. They plan to use it for development of mobile applications for kids.
- A demo was presented at SLT 2014.
- A paper was published at TSD 2015.

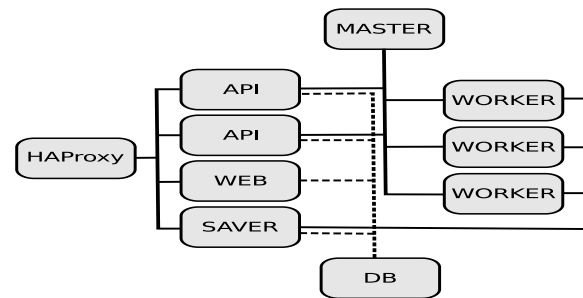
Future Work

- Support for deep neural networks
- Automatic self-learning

Scalability

CloudASR architecture is able to:

- Run on multiple machines
- Scale dynamically
- Handle up to 1000 requests



Customizability

- Kaldi models can be easily used
- Other engines can be integrated by implementing the following interface:

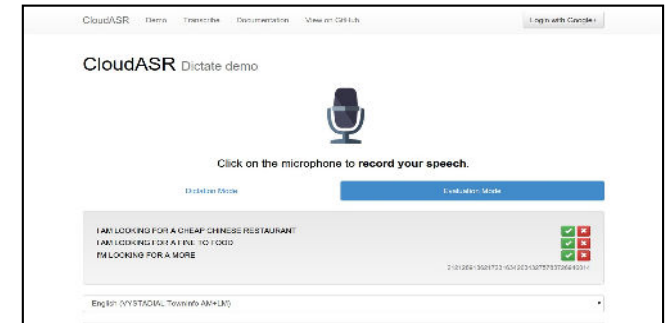
class ASR:

```
def recognize_chunk(self, chunk):
    return (1.0, "Interim hypothesis")
def get_final_hypothesis(self):
    return [(1.0, "Final hypothesis")]
def reset(self):
    pass
```

Easy Deployment

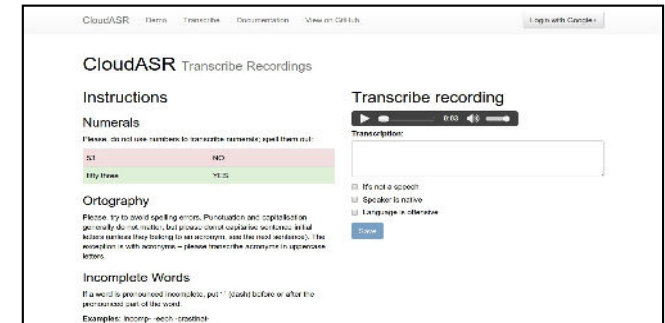
- Can run on single or multiple nodes.
- Only requirement is:
 - **Docker** for a single node deployment
 - **Mesos** for a multiple nodes deployment
- Can be run with single command:
 - **make run_locally** (for single node)
 - **make run_mesos** (for multiple nodes)

CloudASR API



- Batch API is compatible with Google Speech API.
- Online API uses Socket.io library to send chunks and receive interim results.
- Online API can be used in all modern browsers with SpeechRecognition.js library which is a HTML 5 SpeechRecognition drop in replacement.

Annotation Interface



CloudASR contains an annotation interface for transcription of submitted recordings, that can be used for the improvement of ASR systems. Recordings can be transcribed on a CloudASR webpage or at a paid crowdsourcing platform - CrowdFlower.