



Remeeting — Deep Insights to Conversations

Allen Guo, Arlo Faria, Korbinian Riedhammer

Remeeting, Inc.

{allen,arlo,korbinian}@remeeting.com

Abstract

Remeeting is a cloud service that helps you get insights to (spoken) conversations. Audio and video data such as recorded meetings, online conferences, sales or customer success calls are processed using speaker separation and identification, speech recognition and indexing, and an automated keyword analysis. The resulting annotated “documents” can be shared with others and reviewed using a web app that acts as a visual index to the meeting. Furthermore, the extracted metadata is indexed by a search engine to allow for efficient cross-document search. A powerful query DSL allows the user to make sophisticated queries such as “what did X say about topic Y in the first quarter of this year” or “show me the keywords for all meetings where X and Y attended”. Similar to retrieval, a watchdog can be set to deliver real-time insights to operations. Use cases include productivity in meetings, compliance and policies, real time callcenter analytics and better accessibility of large archives.

Remeeting is leveraging, promoting and contributing to open source projects including Kaldi, Elasticsearch and Docker. **Index Terms:** speaker identification, speech recognition, keyword search, big data analytics.

1. Data Ingestion

1.1. Mobile Apps

Remeeting provides mobile voice recorder apps for Android (Play Store) and iOS (App Store) for seamless data upload. Users can record in-person conversations, take notes or timestamps, and take/attach photos. At the end of the recording, or once connectivity is regained, the data is automatically synced and queued for processing.

1.2. Web App

Existing media files or recordings from third party programs such as Citrix GoToMeeting, Cisco WebEx, Zoom or Uberconference can be uploaded on the dashboard of the Remeeting web app.

1.3. API

Developers or third party vendors looking to integrate or facilitate large batch uploads can use the publicly available Remeeting REST API.

2. Data Processing Pipeline

2.1. Storage

Data is stored on secure cloud storage (AWS DynamoDB and S3), including media recordings and metadata such as notes, documents or photos. This storage also contains

user-specific speech processing parameterizations, including speaker-specific acoustic models and topic-adapted language models.

2.2. Speaker Diarization and Linking

Speaker diarization helps to determine who spoke when, resulting in much better insights. A binary key based implementation provides both faster-than-real-time and state-of-the-art accuracy: about 99% for a typical two-person conversation.

2.3. Automatic Speech Recognition

Although accuracy can vary depending on acoustic conditions and content, Remeeting provides “regular” automatic transcripts. A Kaldi-based deep neural network hybrid system using online i-vector based adaptation achieves about 12% WER on *eval2000* at a RTF of about 1.5x, using a vocabulary of about 245k words and a language model with about 100M n-grams.

2.4. Search Indexing

Using custom extensions to Elasticsearch, the Remeeting search index is constructed on top of the *lattice* data structures resulting in a much higher recall when querying the data. Semantic and phonetic query expansion are used to improve the performance and quality of the search.

3. Data Analysis

3.1. API

All data can be queried and retrieved via the Remeeting REST API. Direct query access to the Elasticsearch interface can be provided upon request.

3.2. Remeeting Web App

The Remeeting Web App (see demo) is a platform to manage, review, annotate and share individual recordings, as well as search across the user’s complete data collection (see Fig. 1).

4. Orchestration

Remeeting leverages the open source toolkits Kaldi¹, Elasticsearch² and Docker³. The Remeeting cloud service is architected to be fully scalable, and deployed on the Amazon Web Services cloud.

¹<https://github.com/kaldi-asr/kaldi>

²<https://www.elastic.co>

³<https://www.docker.com>

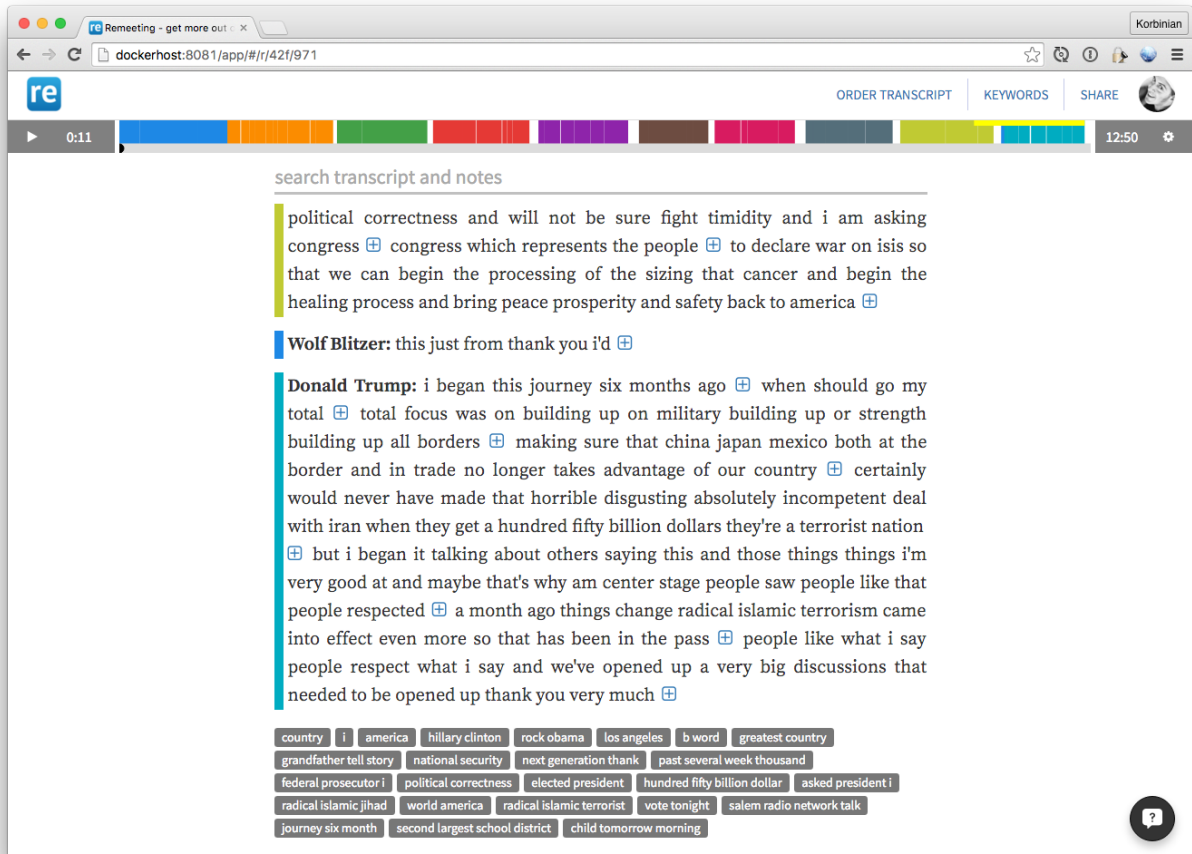


Figure 1: Screenshot of the web app to review and analyze the recording. Users can search for speakers or words within audio and notes, while the color-coded timeline and adjustable scope, see what keywords are mentioned and who said them, and collaborate with note-taking.