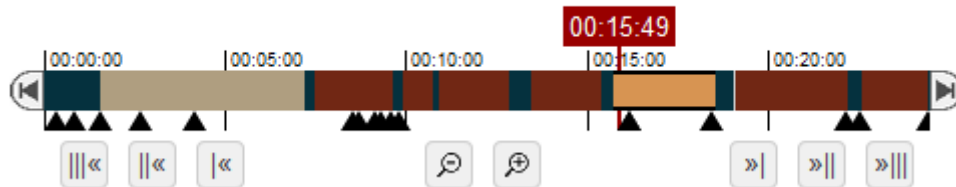


Fraunhofer IAIS Audio Mining „ Solution for Broadcast Archiving“

Dr. Joachim Köhler | LT-Innovate | Brussels | 22.11.2016



☒ Deutschland ☒ Deutschlands



 **Fraunhofer**
IAIS

Outline

- Speech Technology in the Broadcast World
- Deep Learning Speech Technologies
- Fraunhofer Audio Mining System
- Audio Mining Architecture
- Audio Mining User Interface
- Outlook



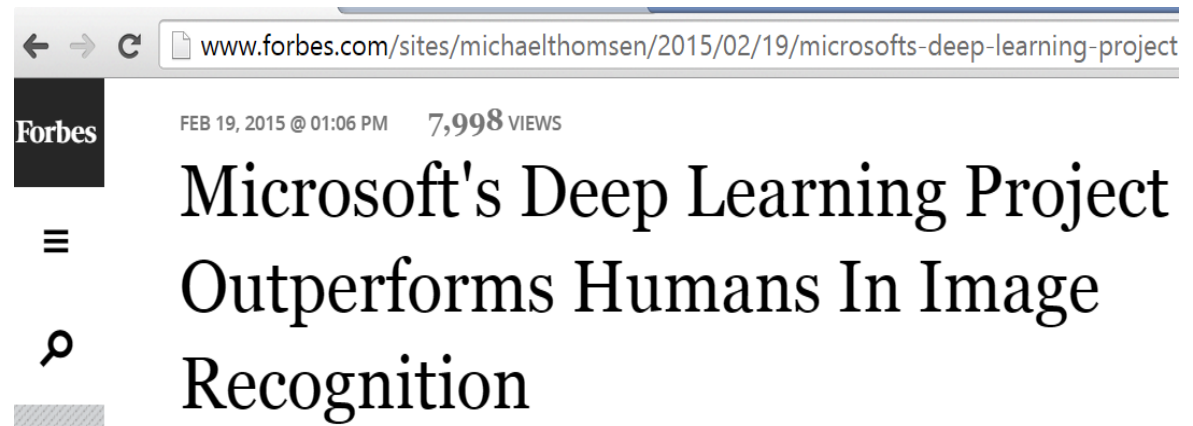
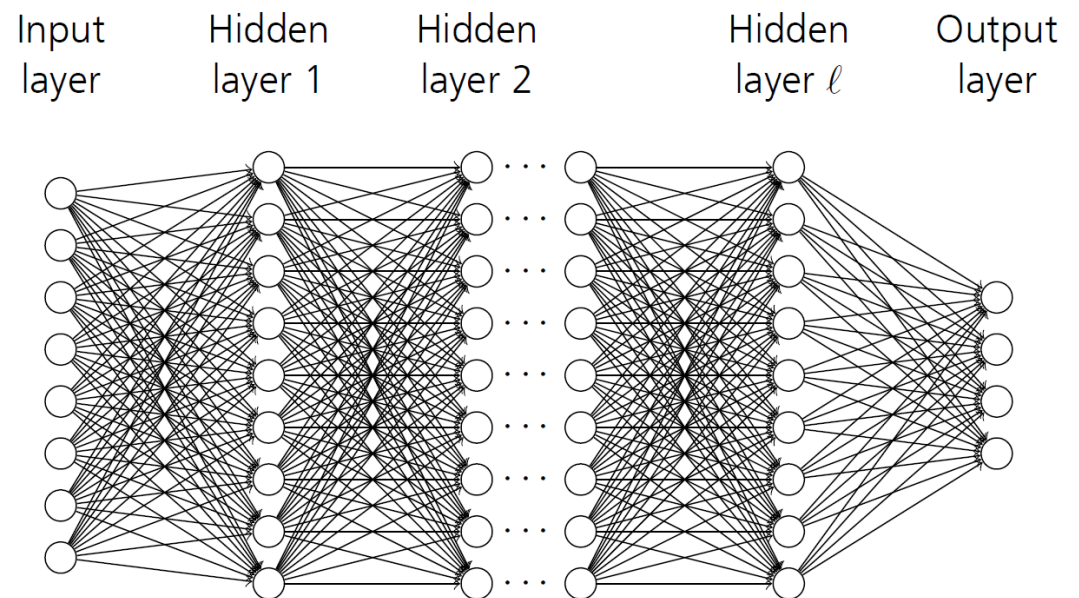
From Research Prototype to Productive Use!

Fraunhofer IAIS Audio Mining

- B2B Speech Recognition Solution for the Media Industry
- Key Facts
 - Large Vocabulary Continuous Speech Recognition (1.000.000 words) optimized for media content
 - Automatic structuring of audio-visual content
- Applications along the Media Asset Chain
 - **Archive:** Indexing and transcription of media archive content
 - **Online:** Search functionalities for media portals (e.g. InClip-Search) and content-based recommendation
 - **Subtitling** of TV content
 - **SocialTV:** Second Screen information enrichment
 - **Advertising:/Marketing** Video Search Engine Optimization (VSEO) and contextualized advertisement

Deep Learning in Speech Recognition for Content Mining

- Massive usage of DNN technologies for pattern recognition and machine learning
- Usage of very large annotated databases
- Significant reduction of word error rates
- Scenarios
 - Image recognition, IBM Watson: Visual Search in AREMA
 - Speech recognition (Siri, Cortana, Google Now, Audio Mining)



Speech Recognition Results and Improvements

- Gaussian Mixture Models (GMM) => Deep Neural Networks (DNN) and Recurrent Neural Networks (RNN)
- Extension of the vocabulary size from 200 000 auf 510.000 words
- Extension of the training database from 105 to 1000 hours for acoustic training
- Extension of the data size for LM training from 40 Mio. words to 1 Mrd. words

Jahr	Acoustic Model	Language Model	Training data [h]	WER [%] planned	WER [%] Spontaneous
2012/13	1. GMM	3gram, 200k	105	26.4	33.5
2013/14	2. GMM	3gram, 200k	323	24.0	31.1
2014	3. DNN	3gram, 200k	323	18.4	22.6
2015	4. DNN	5gram, 510k	1005	13.3	16.5
2016	5. RNN	5gram, 510k	1005	11.9	14.5

Fraunhofer IAIS Audio Mining: Technology

- Speaker diarization to structure recordings automatically (e.g. speaker information)
- ASR System based on KALDI open source package
 - Using Deep Neural Networks
 - Completely speaker independent
 - Real-time processing
- Trained on 1.000 hours large-scale German broadcast database
 - Variety of different TV channels
 - All speech material manually transcribed
- Service-orientated architecture to control and run the recognition engine
- Highly adaptable for specific domains and scenarios



Performance open domain:

- 11,9% word error rate on planned speech
- 14.5% word error rate on spontaneous speech

Speech Recognition for Media Archiving

powered by Fraunhofer IAIS



Customer: WDR, German Broadcaster
(Archive Department)

Project facts:

- Integration of Fraunhofer IAIS Audio-Mining system into the WDR IT environment (ARCHIMEDES und IVZ)
- Content mining of large amounts of AV-data, immediately!
- Better navigation and segmentation of radio and TV material
- Search in spoken utterances
- Full transcription and keyword generation

Technology provided by Fraunhofer:

- Broadcast speech recognition
- Automatic speech segmentation

Speech Recognition

Zeitpunkt: 0:00:12

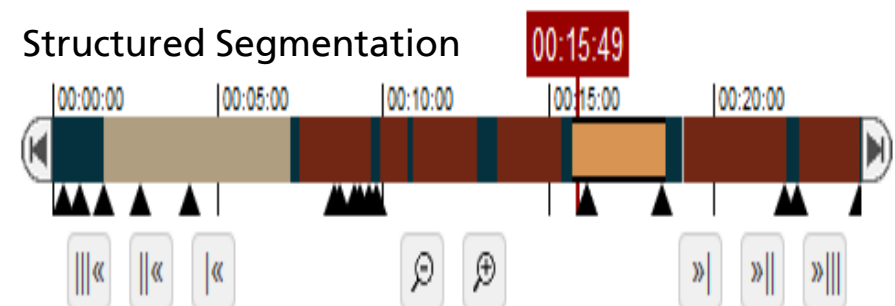
ID 1 ... an Bundeskanzlerin Merkel geschrieben die Überschrift
male **Deutschland** ist ein Überwachungsstaat ...

Zeitpunkt: 0:00:44

ID 1 ... Sommerpressekonferenz haben sie gesagt **Deutschland**
male sei kein Überwachungsstaat seit den ...

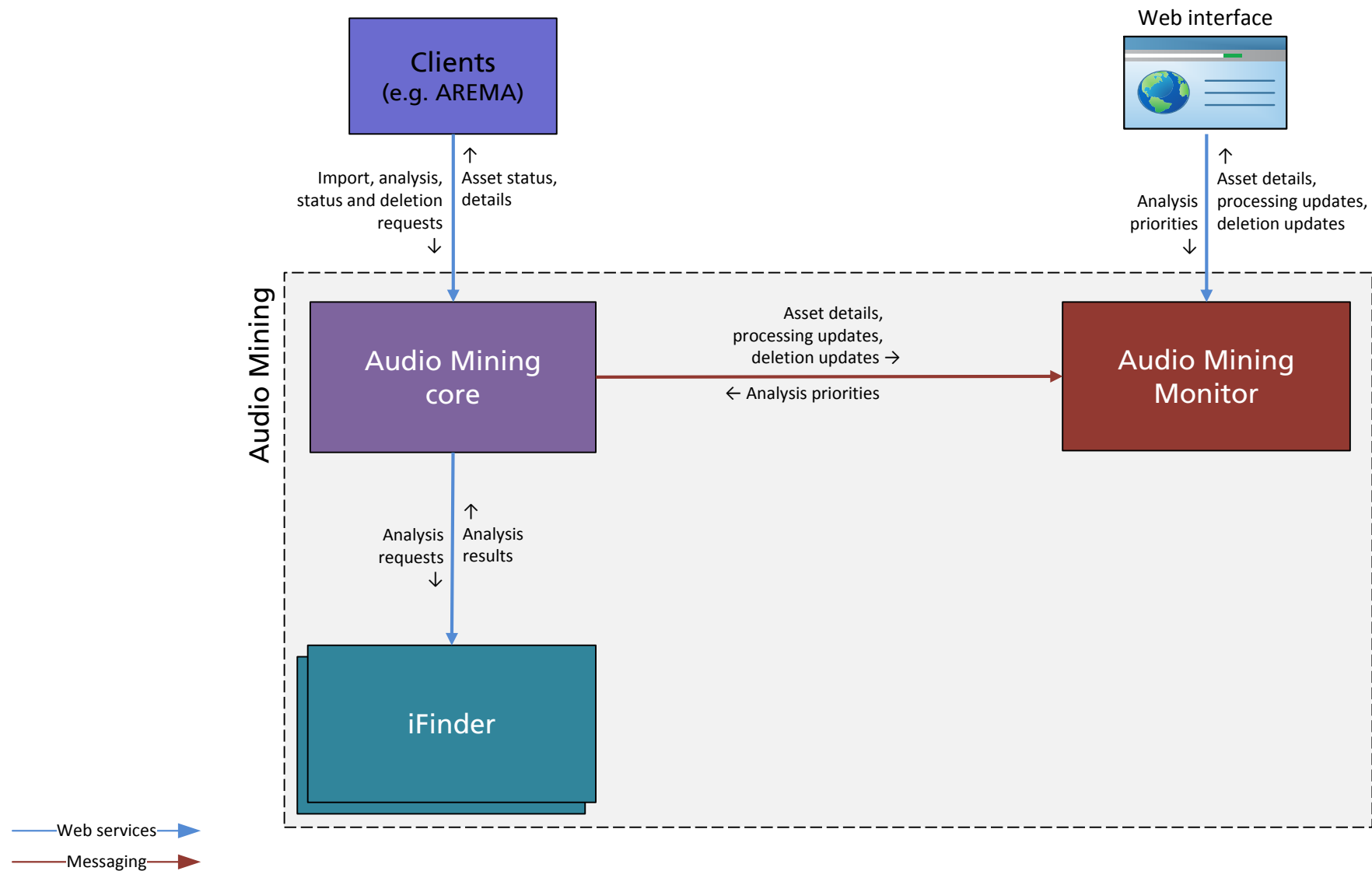
☒ Deutschland ☒ Deutschlands

Structured Segmentation

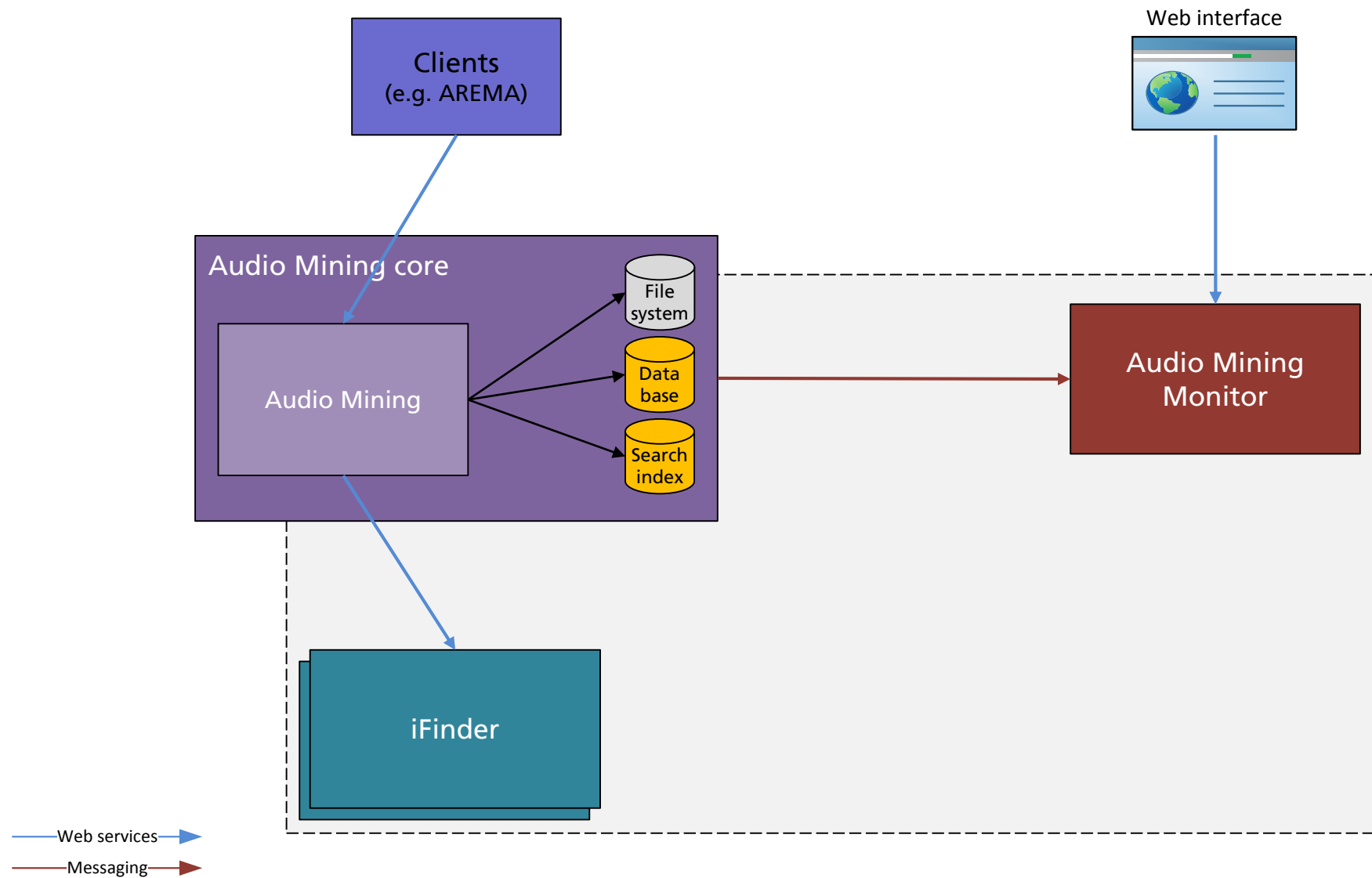


Audio Mining Solution **ARCHITECTURE**

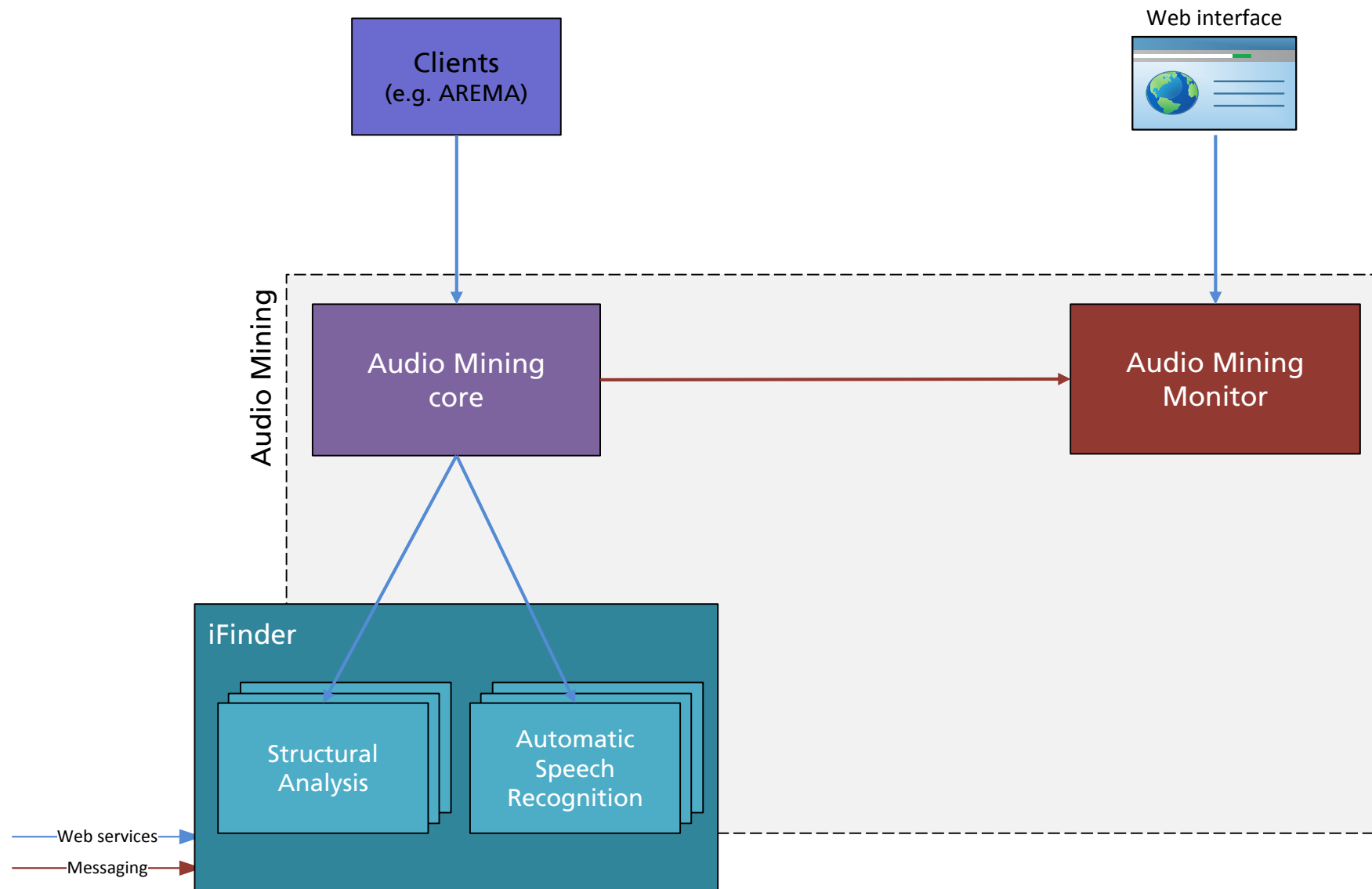
System architecture



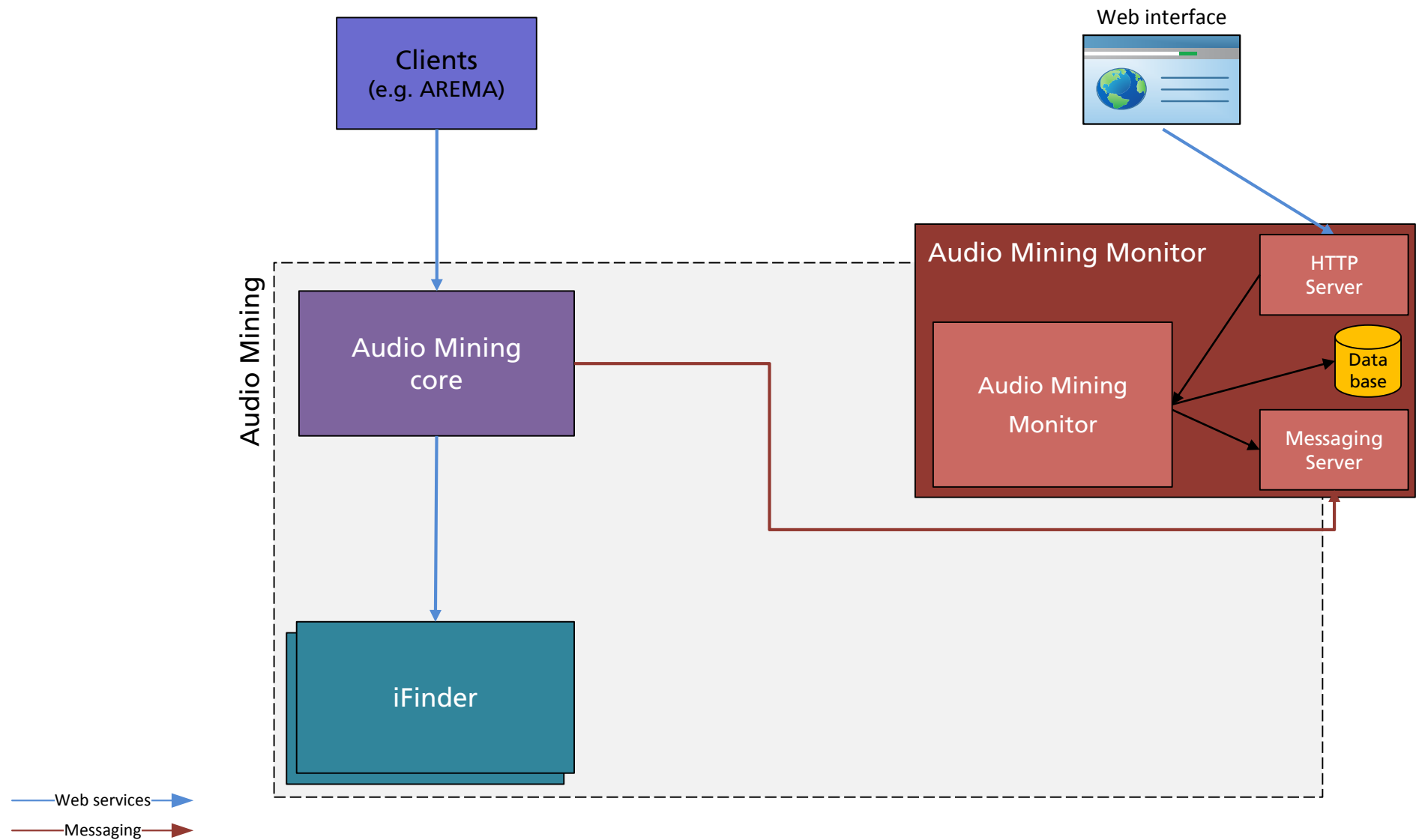
System architecture



System architecture



System architecture



Fraunhofer Audio Mining: Search Interface (I)

<https://nm-prototype.iais.fraunhofer.de/demo/>

AudioMining
powered by **Fraunhofer** IAS

Keywords

Suchen

Leeren

Search functionality:
Find audio and video files with specific keywords, specific words in the title or the transcript, or with a specific series name.

00:00:02

00:00:00

00:00:30

00:01:00

00:01:30

00:02:00

⏮

⏪

⏩

⏭

🔍

⏮

⏪

⏩

⏭

Titel: Ukraine-Krise: Einschätzungen von Ulrich Deppendorf

Sendereihe: Mittagsmagazin

Sender: Das Erste

Sendedatum: 10.2.2015

Dauer: 00:02:10

Keywords:

Gipfel

Minsk

Uli

Deppendorf

Waffenlieferung

Putin

Wladimir

Barack

Vladimir

Francois

Obama

Ansprechpartnerin

Kontaktgruppe

Schlüsselrolle

Player

Überblick | Suchtreffer | Transkript | Empfehlungen

Ukraine-Krise: Einschätzungen von Ulrich Deppendorf

Mittagsmagazin | Das Erste

10.2.2015 | 00:02:10 min.

Gipfel

Minsk

Uli

Deppendorf

Waffenlieferung

Putin

Wladimir

Barack

Vladimir

Francois

Obama

Ansprechpartnerin

Kontaktgruppe

Schlüsselrolle

Player

Sternsinger

N-JOY - Radiokirche bei N-JOY | N-JOY

4.1.2014 | 00:01:40 min.

Berlin

Sternsinger

Pfarrrei

Podcast

Segen

Bindestrich

Hertie

Lubitsch

Luxor

Luc

Naja

Goldberg

Bertin

Unterengadin

Joy

Die Abendschau

Sport in der Abendschau | Bayerisches Fernsehen

7.1.2015 | 00:01:26 min.

Schäden

Segen

Türsteher

Diskothek

Nürnberg

Rückversicherern

Strafmilderung

© Fra

er

Fraunhofer Audio Mining: Search Interface (II)

<https://nm-prototype.iais.fraunhofer.de/demo/>

The screenshot displays the Fraunhofer Audio Mining Search Interface (II). The interface is divided into several sections:

- Header:** The logo "AudioMining powered by Fraunhofer IAI" is on the left. On the right, there are search filters: "Titel" (Title), "Transkript" (Transcript), and "Analysedatum" (Analysis date), each with a dropdown menu and a text input field. Below these are two buttons: "Suchen" (Search) and "Leeren" (Clear).
- Video Player:** A large video player is on the left. It shows a video with a play button in the center. Below the video, there is a timeline with a red bar indicating the current position at 00:00:48. The video has subtitles: "die eigenmächtige Entscheidung Rußlands den Konvoi die ukrainische Grenze passieren zulassen ohne die Genehmigung".
- Search Results:** On the right, there is a list of search results. Each result includes a thumbnail, a title, a date, and a duration. The first result is titled "die eigenmächtige Entscheidung Rußlands den Konvoi die ukrainische Grenze passieren zulassen ohne die Genehmigung" and has a duration of 00:14:41 min. The second result is titled "Traum und Wirklichkeit: Russland und die Modernisierung - 27.08.2011" and has a duration of 00:02:48 min. The third result is titled "Mittagsmagazin | Das Erste" and has a duration of 00:02:48 min.

Annotations:

- Preview functionality:** Select a media file from the right-hand side to watch it or listen to it.
- Subtitles:** Audio Mining creates subtitles based on the transcript and the structural analysis results.
- Segmentation/Speaker clustering:** Audio Mining detects whenever the speaker changes and divides the media file into multiple segments. Jump to a specific segment by clicking the timeline.

Fraunhofer Audio Mining: Search Interface (III)

<https://nm-prototype.iais.fraunhofer.de/demo/>

AudioMining
powered by **Fraunhofer**
IAS

Transkript

Kanzlerin

Transkript

Transkript

Analysedatum

Suchen

Leeren

Überblick | Suchtreffer | Transkript | Empfehlungen

Torsten Sträter: Pressesprecher von Bundeskanzlerin Merkel
extra 3 | Das Erste
28.5.2015 | 00:04:07 min.

Kanzlerin

Neuss

Benedikt

Aussagen

Beige

Marschflugkörper

Eintopf

Frau

Wohnungsbau

Postkarte

Merkel

Verwunderung

NSA

Kokolores

Knows

▼ 7 Treffer

Zeitpunkt: 00:00:09

ID 1 ... Westen dank liebe Bürgerinnen und Bürger die **Kanzlerin** lässt
male herzlich größere aber mit Verwunderung ...

Zeitpunkt: 00:00:31

ID 1 ... Kokolores vielmehr für folgendes zu die **Kanzlerin**
male Bewusstsein sich überhaupt nicht Wohnungsbau der ...

Zeitpunkt: 00:01:13

ID 1 ... X wie natürlich hat Herr Westerwelle die **Kanzlerin** seinerzeit
male dann dahingehend geprüft dass es ...

Zeitpunkt: 00:01:29

ID 1 ... nicht Neuss B sondern Ausbau bei die **Kanzlerin** war dann
male auch umgehend dagegen sagte ...

Zeitpunkt: 00:01:41

ID 1 ... die Union ist niemand überrascht als die **Kanzlerin** selbst dass
male diese amerikanischen Rabauken in ...

Advanced search functionality:
You are also able to search for a specific word inside the transcript.

Word occurrences:
Marks indicate the occurrences of the search term. Click on a mark to jump to the corresponding position in the video.

00:00:08

00:00:00 00:00:30 00:01:00 00:01:30 00:02:00 00:02:30 00:03:00 00:03:30 00

⏮

⏪

⏩

⏭

🔍

⏴

⏵

⏮

⏪

⏩

⏭

Titel: Torsten Sträter: Pressesprecher von Bundeskanzlerin Merkel

Sendereihe: extra 3

Sender: Das Erste

Sendedatum: 28.5.2015

Dauer: 00:04:07

Keywords:

Kanzlerin

Neuss

Benedikt

Aussagen

Beige

Marschflugkörper

Eintopf

Frau

Wohnungsbau

Postkarte

Merkel

Verwunderung

NSA

© Fra

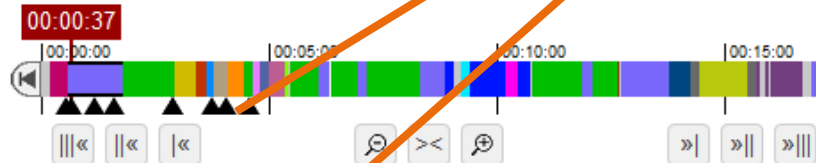
er

Fraunhofer Audio Mining: Search Interface (IV)

<https://nm-prototype.iais.fraunhofer.de/demo/>



☒ Außenminister ☒ Außenminister



Titel: bericht aus berlin

Sendereihe: Bericht aus Berlin

Sender: Das Erste

Sendedatum:
16.1.2011

Dauer: 00:17:16

Keywords:

Außenminister Westerwelle Futtermittel Rettungsschirm Ministerin
Skandal Guido FDP Nationalstaaten Indonesien
Futtermittelhersteller Kubicki Afghanistan Abzug Gutenberg



Letzte Vorbereitungen vor der Eröffnung

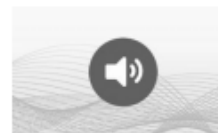
Frankenschau aktuell | Bayerisches Fernsehen
17.11.2010 | 00:02:07 min.

Keywords:

Audio Mining generates keywords for every media file, based on particular relevant words in the transcript.

Again, marks indicate the occurrences of the keyword. Click on a mark to jump to the corresponding position in the video.

Bundestag Grün Niedersachsen



Kindernachrichten 15.06.2013

NDR Info - Kindernachrichten | NDR Info
15.6.2013 | 00:05:49 min.

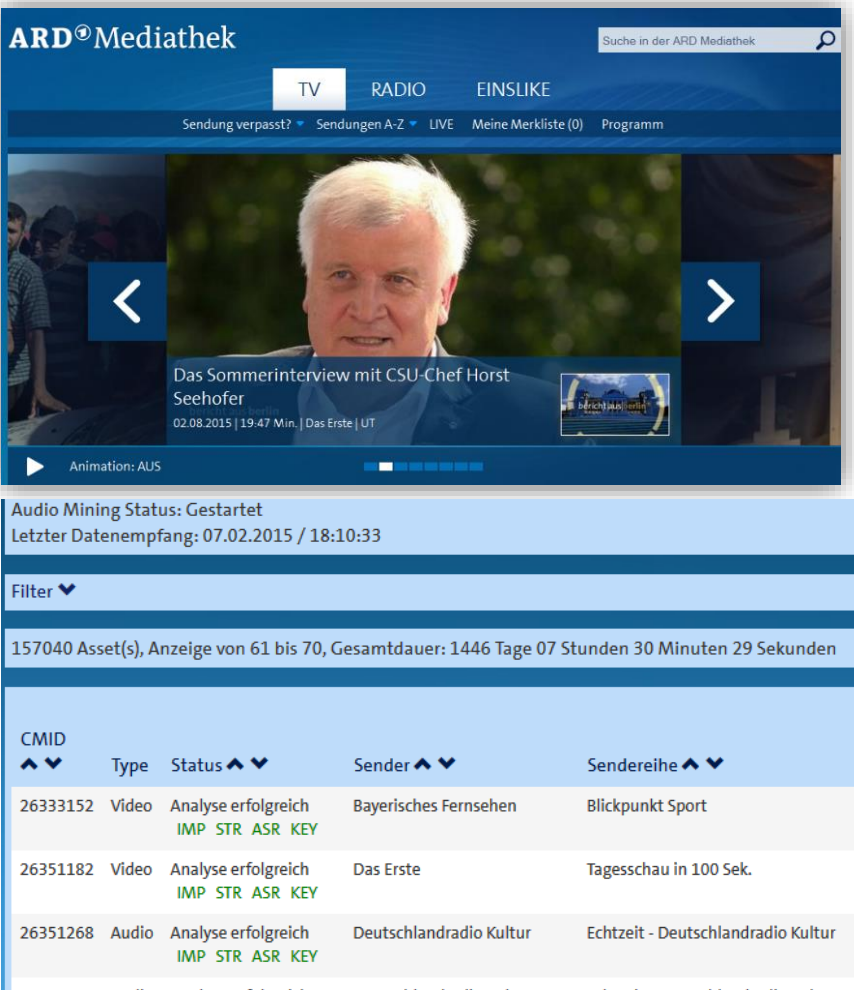
Menschen Gesellschaftsspiel Erdogan
PlayStation Überwachung Geheimdienst
Leute Leiter Türkei Geld Kinderspiel
Vorgehen Jahr Demonstranten Hochwasser

**Die Liberalen und die Dekadenz: Aufstand
der Besservordienenden**

Speech Recognition for ARD Mediathek

Artificial Intelligence application powered by Fraunhofer

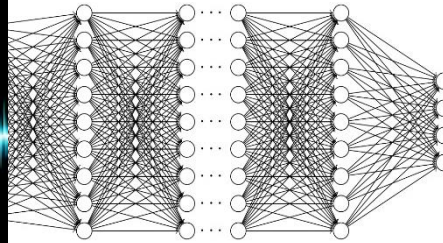
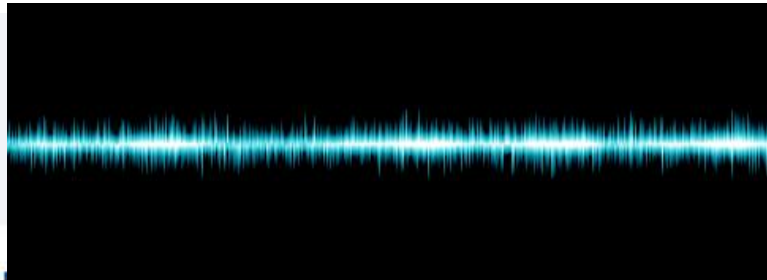
- Content analytics of 200.000 media assets
- Advanced search and retrieval capabilities
- Full transcription of multimedia content
- Daily processing of 2000 new media assets from radio and TV
- Core technology for recommendation and personalization services
- Link: <http://www.ardmediathek.de>



The screenshot displays the ARD Mediathek website interface. At the top, there's a navigation bar with 'TV', 'RADIO', and 'EINSLIKE' tabs. Below this, a search bar and links for 'Sendung verpasst?', 'Sendungen A-Z', 'LIVE', 'Meine Merkliste (0)', and 'Programm' are visible. The main content area features a video player showing a portrait of Horst Seehofer with the title 'Das Sommerinterview mit CSU-Chef Horst Seehofer' and a progress bar. Below the video player, the 'Audio Mining Status' is shown as 'Gestartet' with the last data reception on 07.02.2015 at 18:10:33. A 'Filter' dropdown is present, followed by a summary of 157040 assets. A table lists specific media assets with columns for CMID, Type, Status, Sender, and Sendereihe.

CMID	Type	Status	Sender	Sendereihe
26333152	Video	Analyse erfolgreich IMP STR ASR KEY	Bayerisches Fernsehen	Blickpunkt Sport
26351182	Video	Analyse erfolgreich IMP STR ASR KEY	Das Erste	Tagesschau in 100 Sek.
26351268	Audio	Analyse erfolgreich IMP STR ASR KEY	Deutschlandradio Kultur	Echtzeit - Deutschlandradio Kultur

DNN-based Speech Recognition for Live Subtitling



AudioMining
powered by  **Fraunhofer**
IAIS



Keine Reise in die Ukraine

die frühere Regierungschefin und Symbolfigur der orangenen Revolution sitzt eine siebenjährige Haftstrafe

Outlook

- Audio Mining is accepted in the broadcast domain (Germany)
- Projects in the archive environment are on the way
- First project with large media portals (ARD) are realized
- Future domains in the media (research prototypes)
 - Subtitling (huge amount of TV data is subtitled, either manually or by re-speaking)
 - VSEO
 - Second Screen Applications
- Transfer to other languages and media organisations outside Germany
- Transfer to other domains outside the media world
- Adding more language technologies (TTS, MT, ...)



Contact

Fraunhofer Institut IAIS

Schloss Birlinghoven

53754 St. Augustin

www.iais.fraunhofer.de



Dr. Joachim Köhler

Tel: +49 (0) 2241 14 1900

Mail: joachim.koehler@iais.fraunhofer.de

Link: <https://www.iais.fraunhofer.de/audiomining.html>