LREC 2006 workshop:
Crossing media for improved information access

The iFinder audio-visual indexing framework for cross media applications

23.05.2006

Dr.-Ing. Joachim Köhler

Head of Research of Competence Center NetMedia

http://www.imk.fraunhofer.de

Joachim.Koehler@imk.fraunhofer.de

Outline

1. Introduction
2. Architecture of iFinderSDK
3. Media indexing modules for speech and video
4. Generic iFinder system (MPEG-7 metadata extraction and retrieval system)
5. Project: AGMA (BMBF)
6. Project: Piavida (BMBF)
7. Project: WDR/DW Audiomining
8. Project: Boemie (IST)
9. Project: Live (IST)
10. Summary and outlook on future research issues

iFinder - Introduction

Increasing demand on multimedia indexing tools for several applications areas:

- Broadcast domain
- Archiving of parliament speeches
- Media monitoring
- Internet content (podcasts, videocasts)
- Personal recordings (digital camera, Personal Video Recorders)
- etc…

Technical objectives:

- Providing media indexing solutions for practical usage
- Building real life media indexing solutions

Scientific objectives:

- Support of mid- and highlevel features for multimedia search
- Investigation on multimedia semantics
- High indexing accuracy by optimizing the recognition modules to the target domain
- Domain adaptation (creating and usage of huge corpora)
- Investigation of multimodal fusion approaches

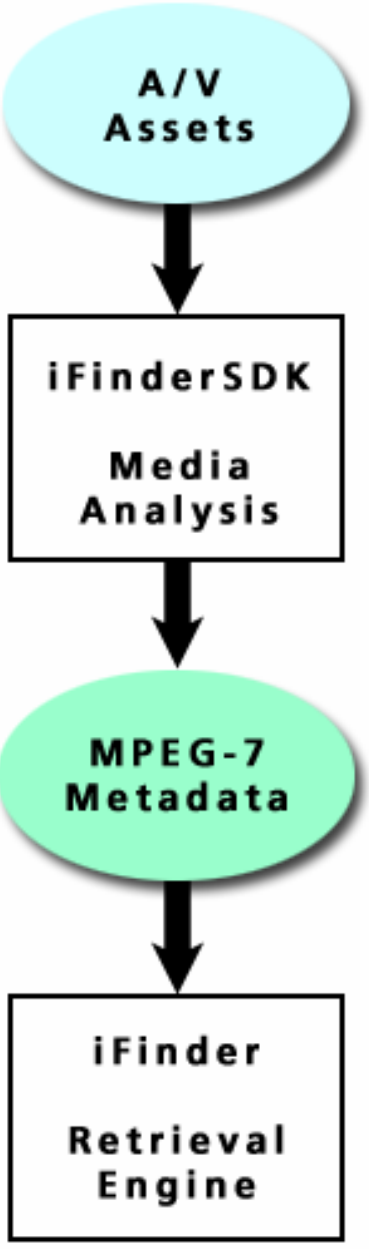
iFinder - Introduction

iFinder framework:

- iFinderSDK (toolbox of multimedia indexing modules)
- iFinder (retrieval application)

Flexible multimedia indexing system:

- Support of common meta data formats (i.e. MPEG-7)
- Extendible media indexing capabilities (integration of new indexing methods)
- Simple and open interfaces to allow reuse and integration of the system
- Should be used and enriched in research projects
- Should be used industrial projects and applications (especially iFinderSDK)



iFinder - History

History:

- BMBF research project AGMA (199 – 2003)
- BMBF research project Piavida (2000 – 2003)
- Definition of iFinder architecture (2002)
- Programming of iFinder framework (2002 – 2005)
- Building several applications and mock ups:
  - IBC 2002, 2004
  - Cebit 2003, 2004
- Prototype for German Parliament
- First industrial project: WDR/DW Audiomining
- IST Project: SHARE (2004 – 2007)
- IST Project: LIVE (2006 – 2009)
- IST Project: Boemie (2006 – 2008)
- IST Project: CitizenMedia (2006 – 2009)
- …

iFinderSDK: Basic Architecture

Metadata and media data are separated streams/files

Currently file based processing

Integration with C++ header files

Common processing interfaces for media Indexing Modules

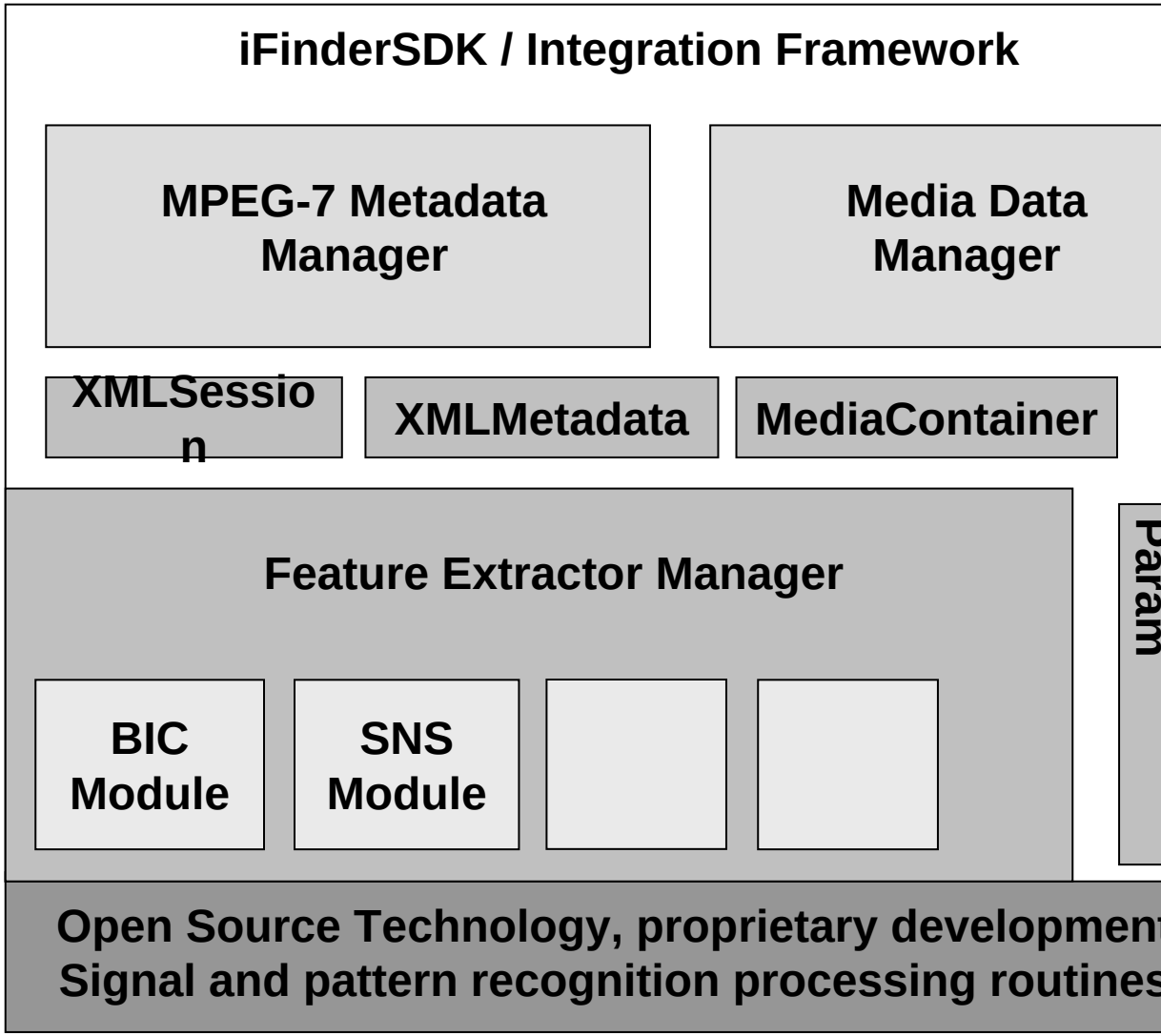
- Init()
- Process()
- Exit()

XML metadata handling:

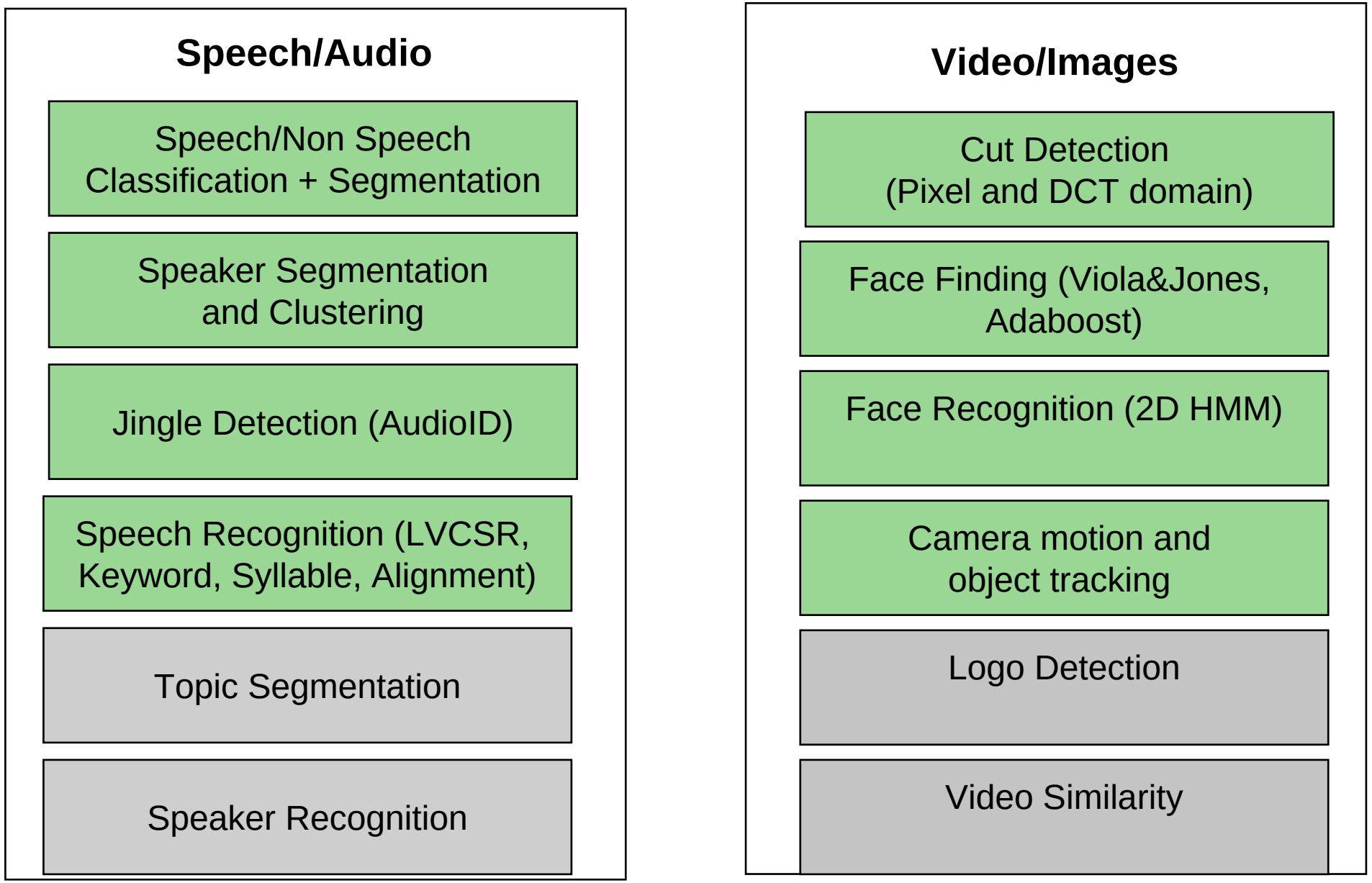
- MPEG-7 descriptions
- Descriptions are enriched step by step
- Xerces parser is used to guarantee MPEG-7 conformance

Parameters:

- Each module can be configured by XML parameter file



Media Analysis: iFinderSDK Toolbox



iFinderSDK: Speech Recognition Module

Basic recognition system: ISIP recognizer

- LVCSR
- MFCC components
- HMM based (Xword triphones)
- Viterbi decoding
- N-Gram language models
- Lattice output

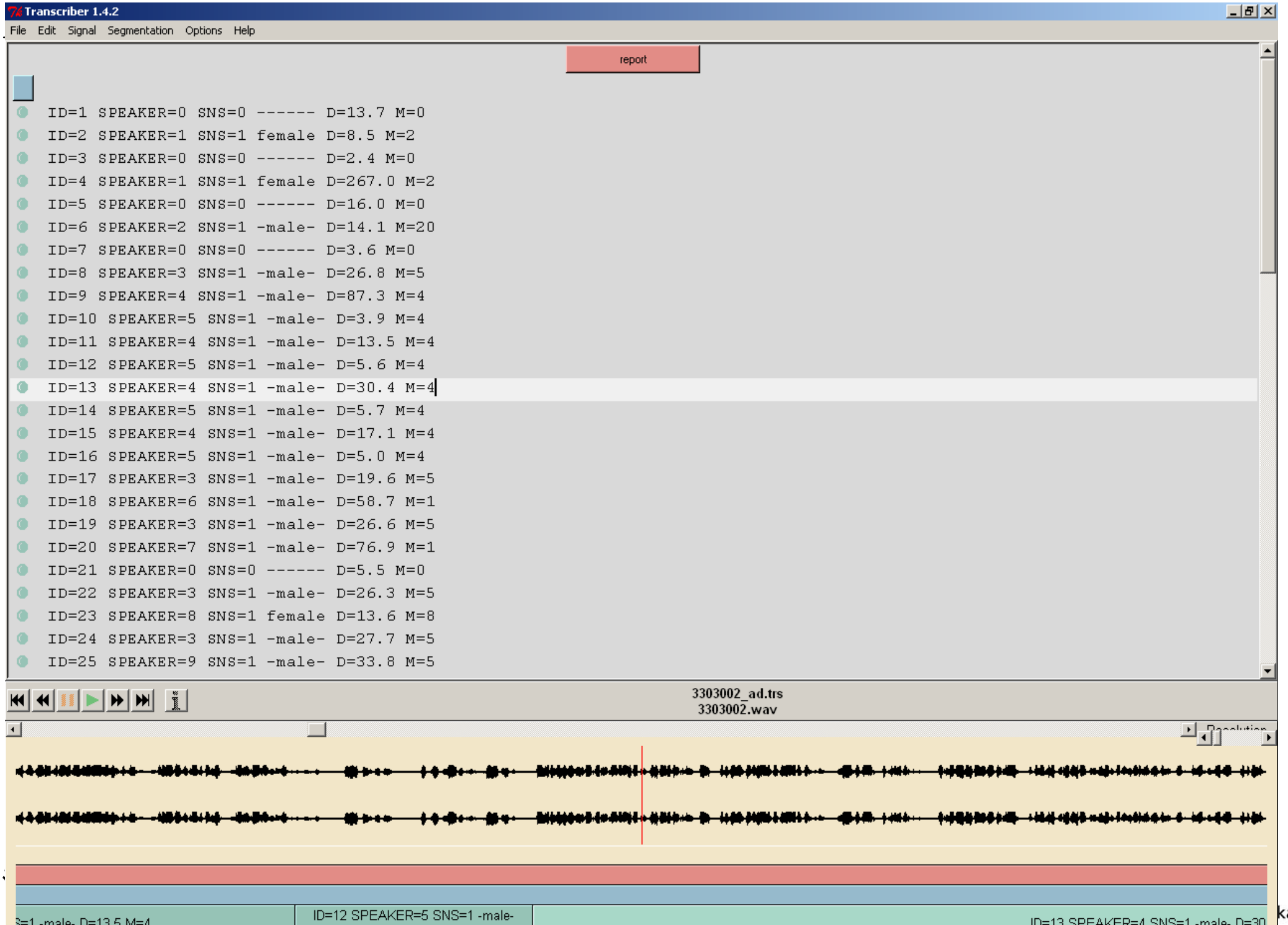
Training of the Acoustic models:

- German broadcast training database (several hours)
- Semi-automatically labeled and annotated
- X-word triphones are trained
- LM trained with SRI toolkit
- To create training database Transcriber is used

Usage of Recognizer

- LVCSR: 100K word recognizer
- Syllable based recognizer (about 5000 syllables for German language)
- Keyword recognition
- Speech text alignment

Usage of Transcriber: Segmentation of Broadcast Signal



iFinderSDK: Text-Speech Alignment

Motivation:

- time accurate access to single words
- Generation of training corpus to retrain HMMs

Two Pass Alignment:

- Processing of large audio files (5-10 hours)
- Pass 1: Recognition of important keywords to find coarse alignment
- Pass 2: Grammar generation for each sentence; recognition with grammar; dynamic programming
- Post-Processing: Calculation of posterior-probabilities to select reliable material

Result:

$$p(w \mid o) = \frac{p'(o \mid w)\,p(w)}{p'(o \mid w)\,p(w) + p'(o \mid w_{alt})\,P(w_{alt})}$$

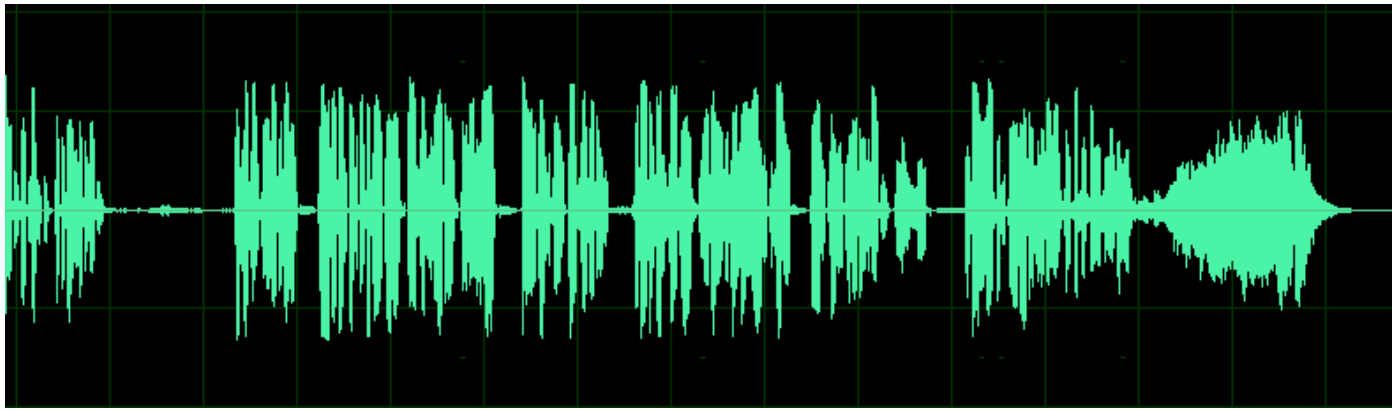
- 20% error reduction with retrained models

Publication:
K. Biatov, J. Köhler: Methods and Tools for Speech Data Acquisition exploiting a Database of German Parliamentary Speeches and Transcripts from the Internet, LREC-2002, Las Palmas, Spain, June 2002

iFinderSDK: Speech Segmentation Methods

Source material:
News show from the broadcaster Deutsche Welle

Wave form file



Segmentation of heterogeneous audio signal:

- Speech/nonSpeech segmentation
- Applause detection
- Jingle detection (HMM based)
- Speaker segmentation based on the Bayesian Information Criterion (BIC)
- Male/Female recognition

Publication:
J. Löffler, K. Biatov, J. Köhler: *Automatic Extraction of MPEG-7 Audio Metadata Using the Media Asset Management System IFinder*, 25th International AES Conference, London, June 2004

iFinderSDK: Speech/nonSpeech Detection Module

Processing Steps:

1. Feature extraction of 12 values based on zero crossing parameters
2. Classification with multivariate Gaussian classifier
3. Optimal smoothing algorithm based on homogeneity calculation (entropy)
   - Find optimal boundaries using dynamic programming
   - Constraints: MinDuration (0,25 second) and MaxDuration(5,0 second)
   - As score value an entropy measurement is calculated:

$$H(s,e) = \sum_{i=s}^{e} m_i \left(\frac{m_i}{N}\right) \log\left(\frac{m_i}{N}\right)$$

L = (N, N, S, S, S, S, N, S, N, N, S)

H =0 H =-0,9

Recognition results for speech frames/segments of German parliament data:

	correct recognized frames using 2	correct recognized segments using 3
speech	76.6%	96.7%

Publication:
K. Biatov, J. Köhler: *An Audio Stream Classification and Optimal Segmentation for Multimedia Applications*, ACM Multimedia, San Francisco, November 2003

iFinderSDK: MPEG-7 Meta Data Generation Process

Audio/Video Data

segmentation,
face finding

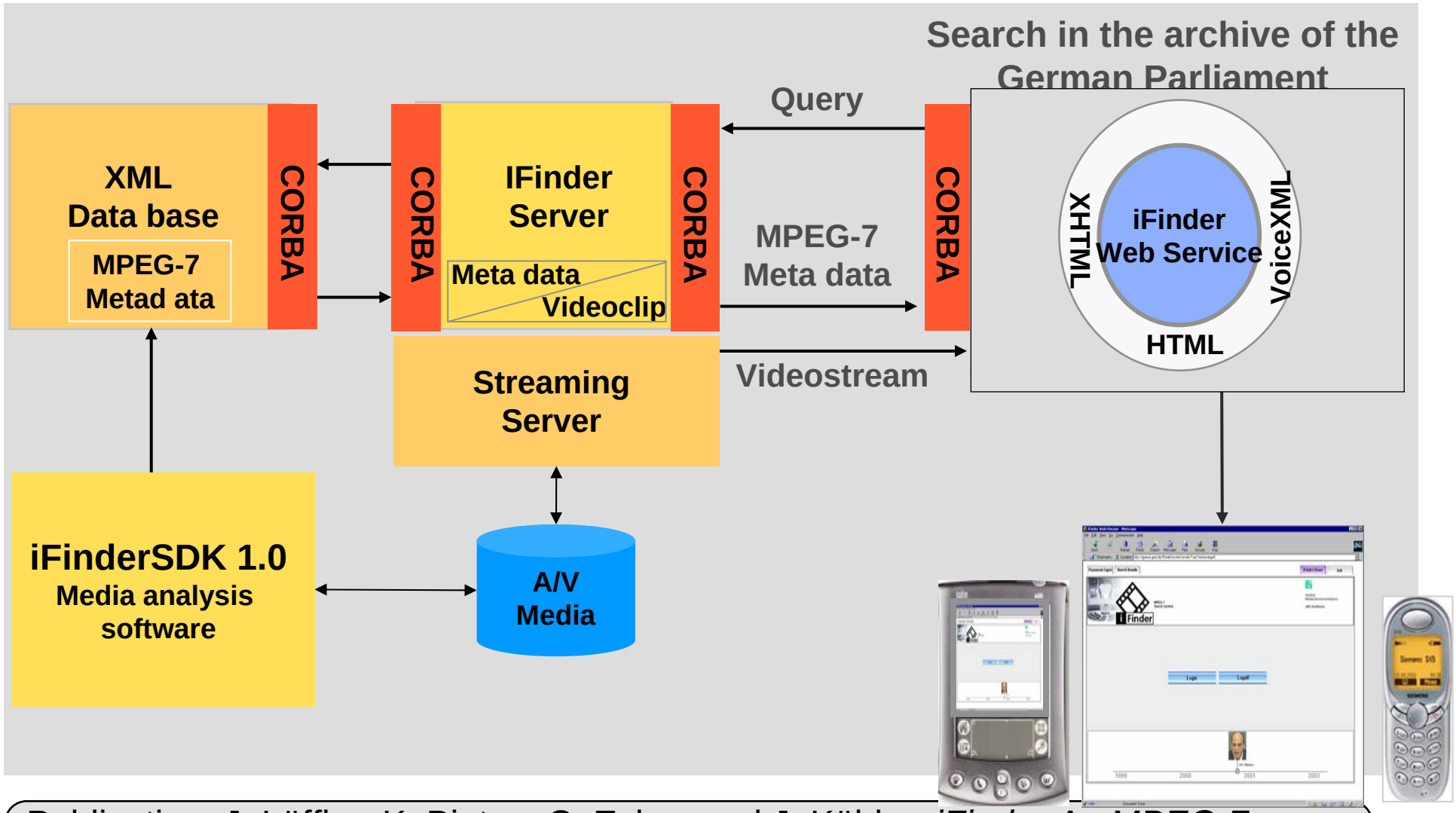
recognition

Stenography:
Die Deutsch
Forschungsfreiheit
….

MPEG-7 XML output

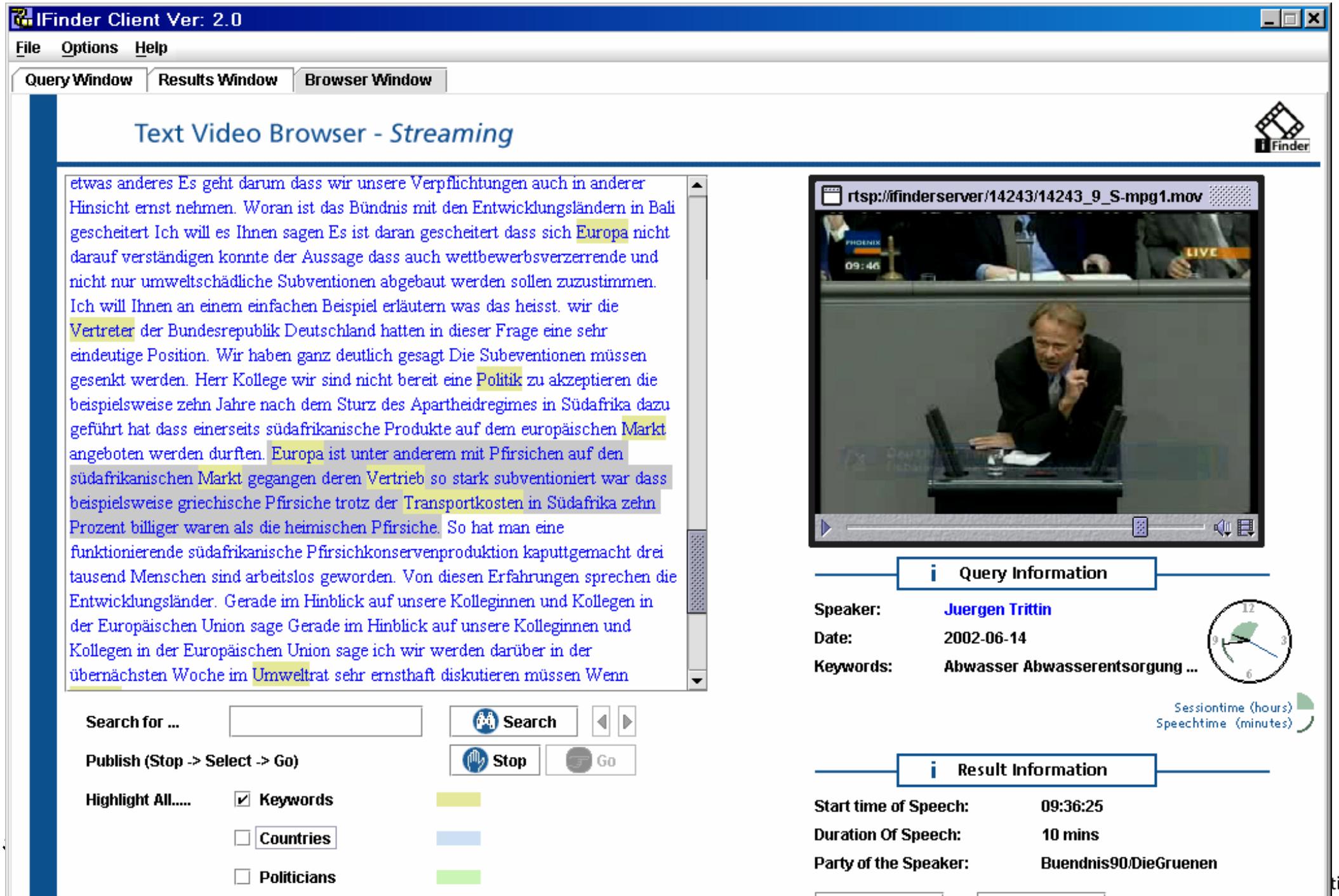
```
<AudioSegment id="Chunk1">
 <TextAnnotation xml:lang="de">
  <KeywordAnnotation xml:lang="de">
   <Keyword>Die Forschungsfreiheit..
  </KeywordAnnotation>
```

iFinder System - Distributed Media Indexing and Retrieval System



Publication: J. Löffler, K. Biatov, C. Eckes und J. Köhler: *iFinder An MPEG-7-based Retrieval System for Distributed Multimedia Content*, ACM Multimedia, Juan Les Pins, France, December 2002

iFinder: Retrieval User Interfaces (MPEG-7 based Text Video Browser



Research Project AGMA (Automatic Generation of Metadata based on MPEG-7)

Project facts

- BMBF project (1999 – 2003)
- Support and exploitation of MPEG-7
- Speech and video indexing methods

Research Goals:

- Multimodal indexing (speech and video features)
- Automatic exploitation of very large training databases

Data and domain:

- Recordings of the German Parliament (> 100 hours)
- MPEG-2 streams from German broadcaster Phoenix
- Closed speaker set (>500 speakers)
- Transcripts are available as stenography
- Face database was collected (manually selected and annotated)

Research Project AGMA (Automatic Generation of Metadata based on MPEG-7)

iFinderSDK modules:

- Speech recognition (speech text alignment)
- Speaker segmentation
- Speech/Nonspeech detection (applause detection)
- Cut detection
- Face detection and recognition (50 faces)
- Person segmentation

Retrieval application

- Xindice database
- Client server application based on Corba
- Java frontend
- Web services to access content over Web technogies (browser, VoiceXML)

Research Project Piavida (Personalized Interactive Audio, Voice and Video Information Portals and Applications)

Project facts

- BMBF project (2000 – 2003)
- Speech and video indexing methods
- Mobile data access

Research Goals:

- Dataminig on multimedia data (spoken document retrieval)
- Multimodal indexing (speech and video features)

Data and domain:

- Recordings from the broadcaster DW (80 hours of “Kalenderblatt”)
- Realstreams are recorded and decoded to PCM data
- Transcripts are available and were used to build training database

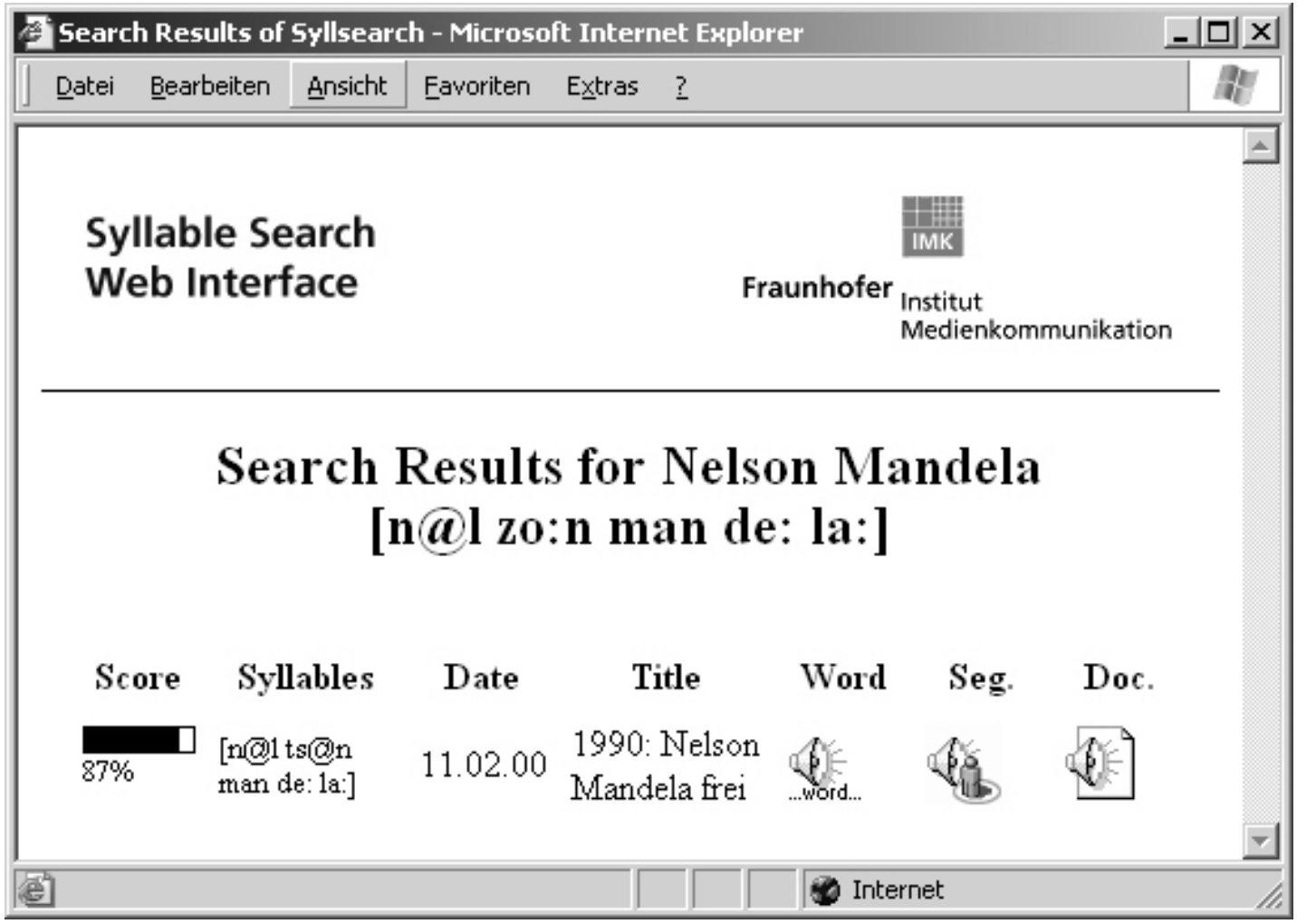
Research Project Piavida (Personalized Interactive Audio, Voice and Video Information Portals and Applications)

iFinderSDK modules:

- Speech recognition
  - speech text alignment
  - Syllable based recognition
- Speaker segmentation

Retrieval application:

- Web based demo application



Project Audioming (with broadcaster Deutsche Welle and WDR)

Project facts

- Industrial project
- Development of an pilot application for audiomining
- 2005/06

Research and Development Goals:

- Automatic generation of a structure of broadcast recordings
- Indexing and search for word queries (e.g. “Angela Merkel”)
- Standalone system

Data and domain:

- WDR and DW provided 160 hours of broadcast recordings (4 different types of radio shows)
- Several hours are transcribed and labelled
- Mixture of speaking styles:
  - Professional and unprofessional speakers
  - Studio and telephone speech
  - Over lapped by music (background music)
  - Dialogue based speaking style versus read speech

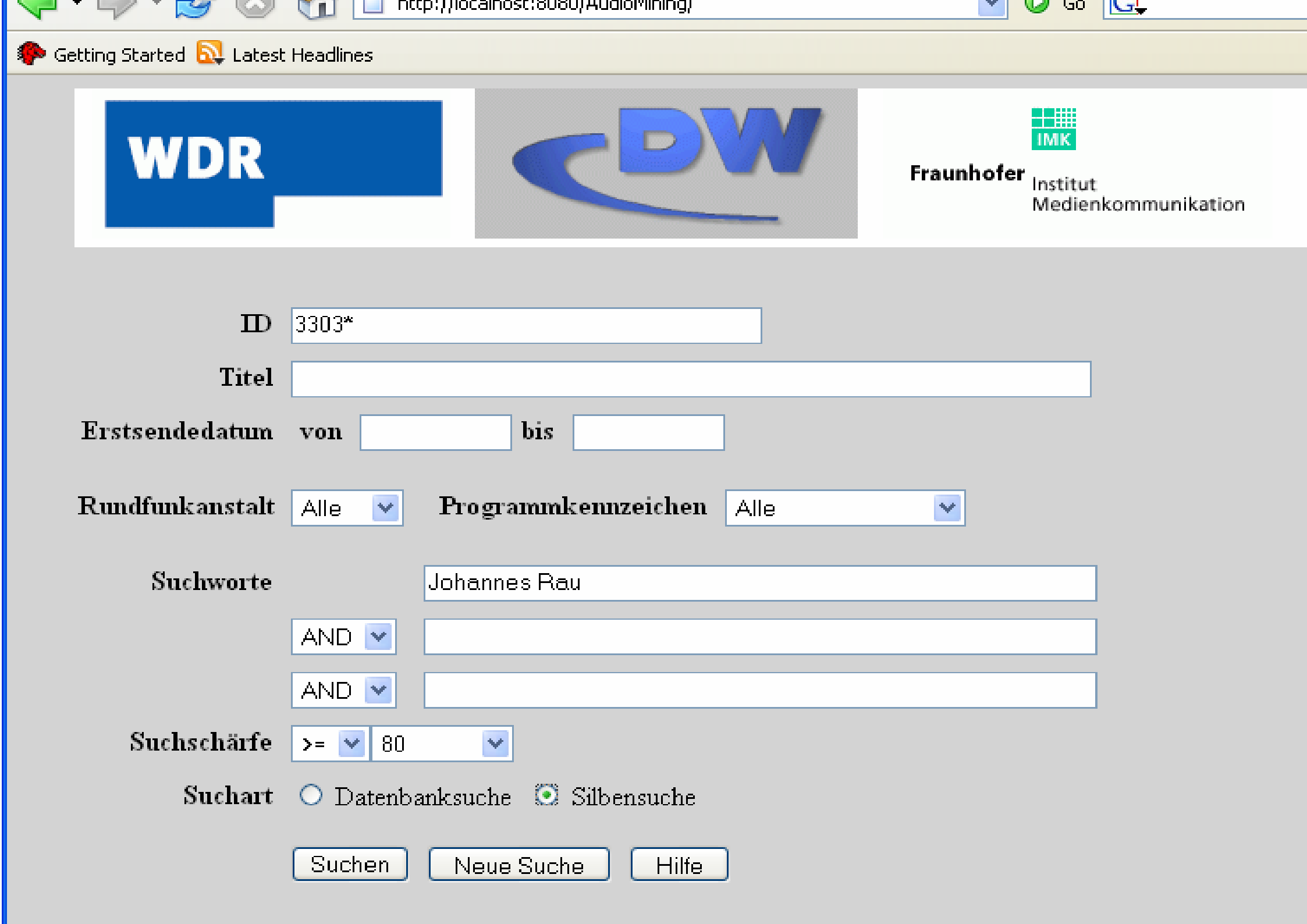
Project Audioming (with broadcaster Deutsche Welle and WDR)

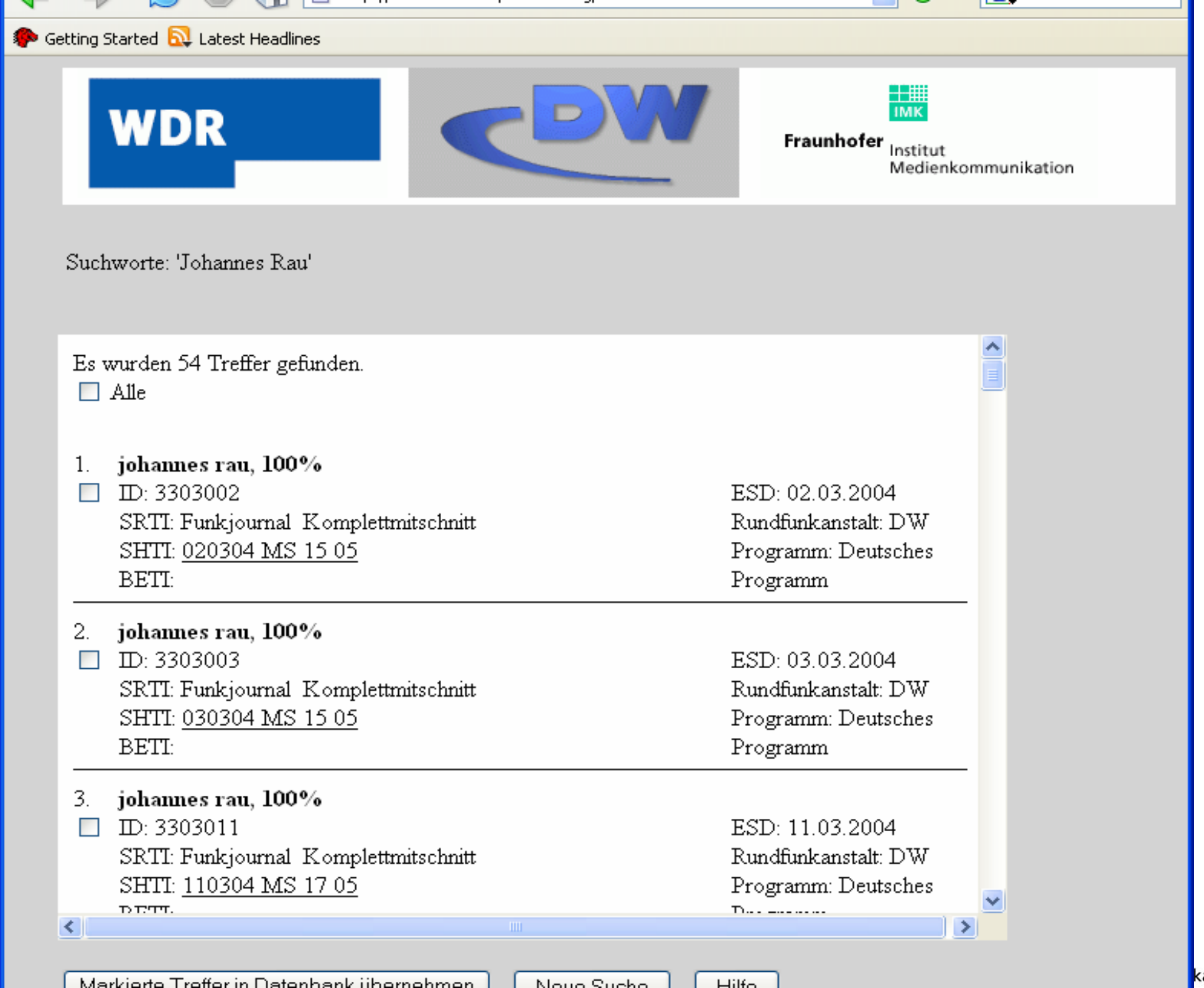
iFinderSDK modules:

- Speech recognition, syllable based recognition (using the Piavida models)
- Speaker segmentation and clustering

Retrieval application

- Content based access to audio segments based on syllable based ASR
- Graphical presentation of automatically generated metadata
- Web based system with Flash Audio browser and full workflow integration
- JAVA middleware (i.e. Java servlets)
- Installation on a standard PC


http://localhost:8080/AudioMining/
Go
Getting Started
Latest Headlines
WDR
DW
IMK
Fraunhofer
Institut
Medienkommunikation
ID
3303*
Titel
Erstsendedatum
von
bis
Rundfunkanstalt
Alle
Programmkennzeichen
Alle
Suchworte
Johannes Rau
AND
AND
Suchschärfe
>=
80
Suchart
Datenbanksuche
Silbensuche
Suchen
Neue Suche
Hilfe


Getting Started
Latest Headlines
WDR
DW
Fraunhofer
IMK
Institut
Medienkommunikation
Suchworte: 'Johannes Rau'
Es wurden 54 Treffer gefunden.
Alle
1. johannes rau, 100%
ID: 3303002
SRTI: Funkjournal Komplettmitschnitt
SHTI: 020304 MS 15 05
BETI:
ESD: 02.03.2004
Rundfunkanstalt: DW
Programm: Deutsches Programm
2. johannes rau, 100%
ID: 3303003
SRTI: Funkjournal Komplettmitschnitt
SHTI: 030304 MS 15 05
BETI:
ESD: 03.03.2004
Rundfunkanstalt: DW
Programm: Deutsches Programm
3. johannes rau, 100%
ID: 3303011
SRTI: Funkjournal Komplettmitschnitt
SHTI: 110304 MS 17 05
ESD: 11.03.2004
Rundfunkanstalt: DW
Programm: Deutsches
Markierte Treffer in Datenbank übernehmen
Neue Suche
Hilfe

ID: 3303002
BETI:
SRTI: Funkjournal Komplettmitschnitt
SHTI: 020304 MS 15 05
Rundfunkanstalt: DW, Programmkennung: Deutsches Programm, Erstsendedatum: 02.03.2004
Suchworte: 'Johannes Rau'

1 Treffer

Start	Dauer	Segmentart	Sprecher
00:10:26	00:01:13	speech	Speaker_M_ID-7
00:11:39	00:00:03	speech	Speaker_M_ID-3
00:11:43	00:00:05	nonspeech	
00:11:49	00:00:25	speech	Speaker_M_ID-3
00:11:53	johannes rau 100%		
00:12:14	00:00:14	speech	Speaker_F_ID-8
00:12:28	00:00:27	speech	Speaker_M_ID-3
00:12:56	00:00:33	speech	Speaker_M_ID-9
00:13:30	00:00:05	speech	Speaker_M_ID-3
00:13:35	00:00:47	speech	Speaker_M_ID-9
00:14:23	00:00:11	speech	Speaker_M_ID-3
00:14:35	00:00:54	speech	Speaker_M_ID-9
00:15:29	00:00:14	speech	Speaker_M_ID-3
00:15:43	00:00:24	speech	Speaker_M_ID-9
00:16:08	00:00:07	speech	Speaker_M_ID-3
00:16:16	00:00:17	speech	Speaker_M_ID-9
00:16:34	00:00:34	speech	Speaker_M_ID-3
00:17:08	00:00:10	speech	Speaker_F_ID-8
00:17:19	00:00:16	speech	Speaker_M_ID-10
00:17:36	00:01:04	speech	Speaker_F_ID-8
00:18:41	00:00:14	speech	Speaker_M_ID-10
00:18:55	00:00:07	speech	Speaker_F_ID-8
00:19:02	00:00:03	speech	Speaker_M_ID-4

> □ || >> << >>> <<< >| |<

Vorlauf [sec] 0

Lautstärke 10

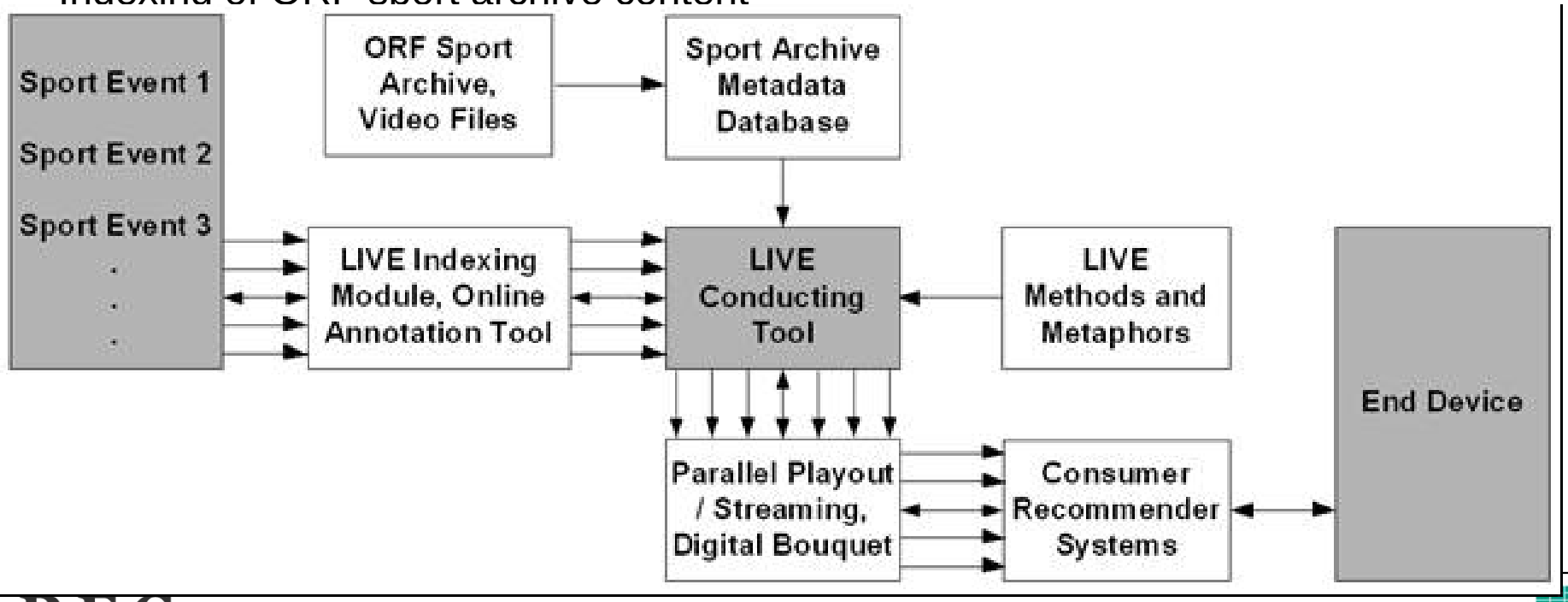
EU IST Project LIVE (SO: 2.4.7)

Project facts

- IST project (2006 – 2009)
- Overall project goal: new iTV production methods for live sport content
- IP project (11 Mio. Budget)
- Coordinator: Fraunhofer IMK

Research Goals on metadata extraction:

- Indexing methods for live content
- Semi-automatic annotation tool for live content
- Indexing of ORF sport archive content



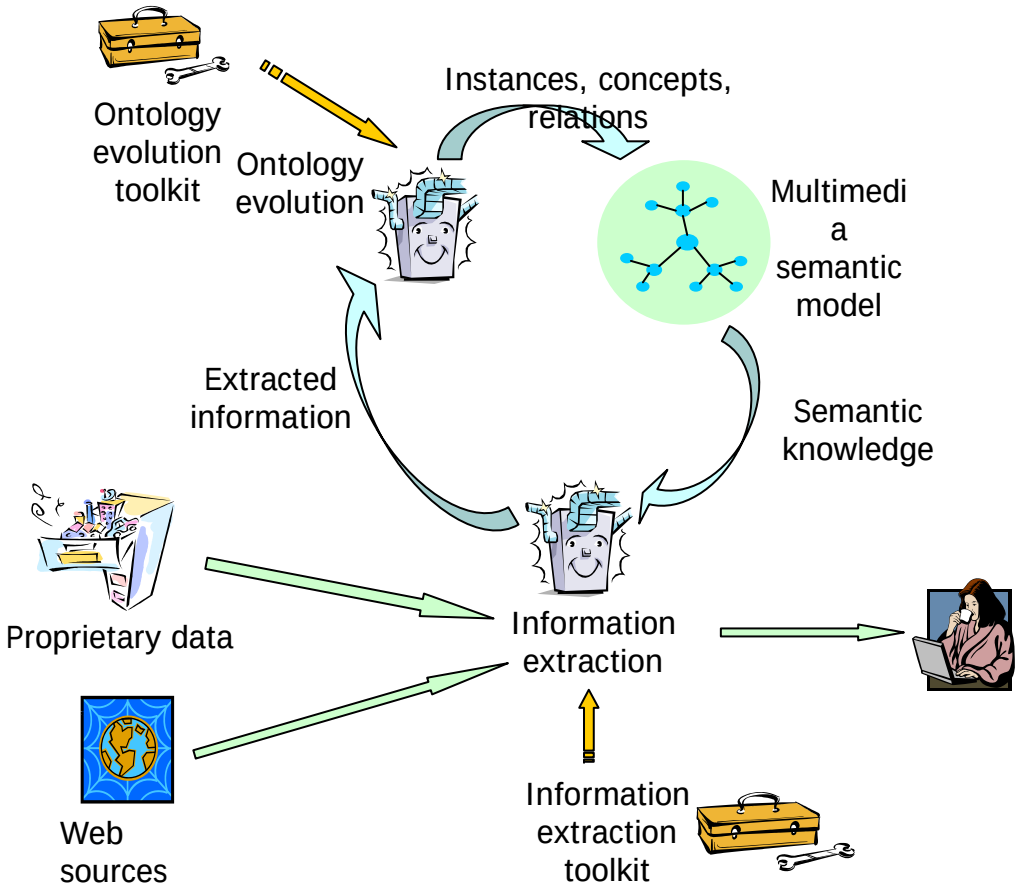
EU IST Project BOEMIE (SO: 2.4.7)

Project facts

- IST project (2006 – 2008)
- Overall project goal: research on multimedia ontologies and semantics
- STREP project
- Coordinator: NCSR

Research Goals on metadata extraction:

- New audio and video indexing methods
- Semantic models for audio visual data
- sport domain



Summary and Outlook on Future Research Issues

- iFinderSDK provides a set of modules for media indexing
- iFinder is used in several research and development projects
- Research issues:
  - Indexing of live content
  - Semantic multimedia indexing
  - High accuracy indexing for each module
  - Semi-automatic generation of annotated databases
- Other projects:
  - EU IST SHARE
  - EU IST CitizenMedia