

MEDIA MINING INDEXER

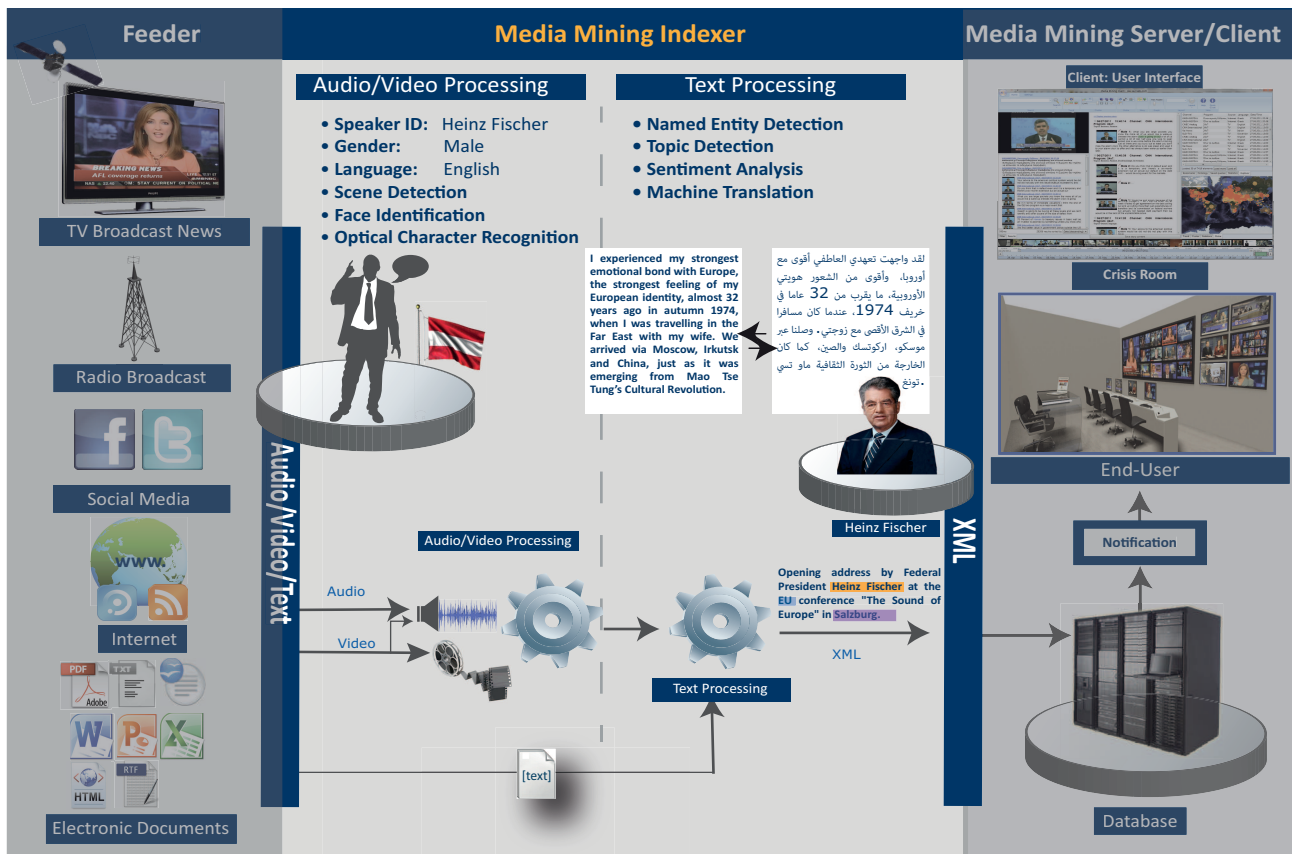
From Unstructured Audio/Video Content to Actionable Media Intelligence

The Media Mining Indexer represents “the heart” of the Media Mining System. Here, speech recognition and other technologies take place i.e. speech becomes text.

The Media Mining Indexer receives media streams through media feeders. These convert input from sources such as radio and TV channels, internet, press agency news feeds, archived media streams, etc. to a format compatible as input for the Media Mining Indexer. The Media Mining Indexer converts this input into a rich-transcript XML describing the content and metadata about the content. This rich-transcript is then sent to the Media Mining Server for storage, search and retrieval.

Some of the exciting Features of the Media Mining Indexer include

- Automatic Speech Recognition enabling accurate conversion of speech to text, even in the presence of high levels of background noise
- Speaker Change Detection segregating audio segments based on speaker and audio environment for focused information retrieval
- Speaker Identification attributing statements to certain speakers ("who" is saying "what" in the news?)
- Gender Identification
- Topic Classification helping to locate stories relating to user-specific interests
- Named Entity Detection tagging words with categories of entities such as persons, locations, organizations.
- Real-time Indexing: Transcripts are generated in parallel to the broadcast.
- Language Model Toolkit: The Language Model Toolkit enables addition of domain-specific vocabulary and context to help keep pace with a rapidly changing news environment.



Features List

License Terms

- Flexible license (scheduled recording multiple channels/languages)
- Unrestricted channel switching
- Unrestricted language change (for all licensed languages)

Automatic Speech Recognition

- State-of-the-art accuracy
- Phonetic vocabulary
- Word spotting
- N-best hypothesis generation
- Audio processing capabilities (command-line interface)
- API support for C++, plain C
- Microsoft Visual Studio integration
- Real-time recognition
- Unlimited vocabulary size
- 50.000+ words base vocabulary
- Speaker/audio segmentation
- Speaker/audio classification
- Language Model Toolkit (LMT): modify base vocabulary
- Language Model Toolkit (LMT): build own vocabulary

Supported Languages

- US English
- International English
- Spanish
- French
- German
- Modern Std. Arabic
- Russian
- Polish
- Greek
- Norwegian
- Hebrew
- Catalan
- Italian
- Mandarin
- Persian (Farsi)

Video Analysis

- Intelligent Keyframing
- Dynamic Keyframes (keyframe server)
- Story-line keyframes (keyframe bar)
- Facial identification
- Optical Character Recognition (screen-OCR)
- Scene Analysis (Scene Change Detection, auto-clipping)
- Entity Extraction

Audio Analysis

- Topic detection and categorization
- Linguistic origin of speaker (Language/Dialect ID)
- Speech/non-speech detection
- Signal-to-Noise Ratio calculation
- Over-/Undermodulation warning
- Music/Jingle/Melody detection and identification
- Named-entity detection
- Flexible named entities (user editable named entities tables)
- Flexible named entity types (add to base: person, location, organization)

Hardware Prerequisites

- INTEL Core 2 Duo E4500 2.0 GHz (or better)
- 2GB-DDR 2 RAM (on top of OS)
- Disk space: 60 GB recommended for production use

Software Prerequisites

- Windows XP Prof (SP3)
- Windows 2003 Standard
- Windows Vista
- Windows 7 (32 bit or 64 bit)
- Windows 2008 Server
- Windows 2008 R2 Server
- ActivePerl-5.12.4.1205 (32bit) from ActiveState
- Linux