

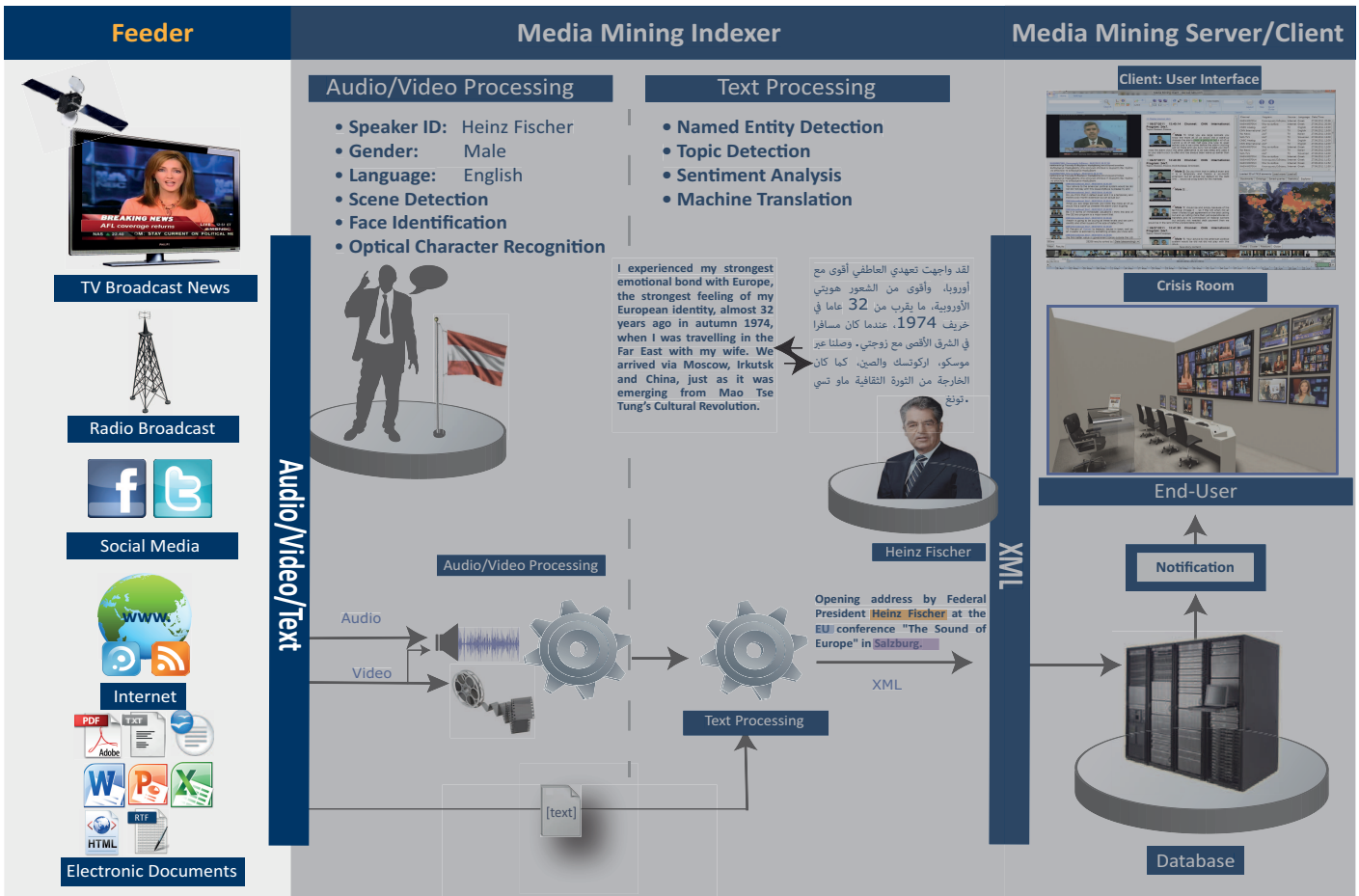
MEDIA MINING FEEDER

for TV & FM Radio, Electronic Documents and Web

Unstructured data from multiple sources (digital satellite TV and radio, FM radio, podcasts, streaming media) is processed into structured, searchable and easily accessible information. During this process video and audio signals are transcoded by the Media Mining Feeder and uploaded onto the Media Storage. Furthermore, the Media Mining Feeder sends audio and meta-data to the speech recognition engine, i.e. the Media Mining Indexer. Currently Media Mining Feeder software supports feeding from the following types of signal source:

- DVB-S (digital TV and Radio broadcast over satellite)
- FM-Radio (analog FM Radio)
- Analog TV
- Capture (analog Audio/Video capture input)
- Local media files
- IP-TV
- WebCollector (gathers data from the Internet)
- Twitter Collector
- Feeds (RSS & Atom)
- Electronic Documents

The Sail Labs Media Mining Universe



Features List

License Terms

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

Supported Ingestion Sources

- External Audio/Video Win7 compliant tuning device (analog)
- Internal Audio/Video analog decoder-module
- Internal DVB-S (digital SD) decoder-module
- Internal DVB-S2 (digital HD) decoder-module
- Internal DVB-T (digital terrestrial) decoder-module
- Internal DVB-C (digital Cable-TV) decoder-module
- Internal satellite radio decoder-module

Supported File Ingestion Formats (Audio/Video)

- MS Windows XP codecs
- MS Windows 7 codecs (WMV9, MPEG2, DIVX, H.264 video and AAC)
- Apple QuickTime (.mov, .qt)
- Real Networks (.rm, .rmvb)
- Windows Media Format (.wmv)
- Windows Media Format Video Codec-1 (VC-1, .wmv9)
- Motion Picture Expert Group (.mpeg, .mp2, .mp4)
- Flash Video (.flv)
- DivX/Xvid
- Ogg/Theora (.ogg, .ogv)
- Matroska container format (.mka, .mkv)
- Transport Stream (.ts)

Electronic Documents Ingestion

Web:

- Web crawling
- Language identification (for multilingual web pages)
- Semantic key story content extraction (link/story segmentation)
- Ad & link suppression
- Selective include/exclude patterns
- Offline mode (with web page snapshot in pdf format)
- RSS & Atom support
- Social Network support (Facebook, IMDb, LinkedIn)
- Autologin (restricted websites access)
- Twitter support

Email:

- pop3 Server support
- IMAP server support

File:

- Plain Text (.txt)
- Microsoft Word (.doc, .docx)
- Microsoft PowerPoint (.ppt, .pptx)
- Microsoft Excel (.xls, .xlsx)
- Portable Document Format (.pdf)
- Open Document Text (.odt)
- Rich Text Format (.rtf)
- HyperText Markup Language (.html)

Hard- and Software Prerequisites

- INTEL Core 2 Duo E4500 2.0 GHz (or better)
- 2GB-DDR 2 RAM (on top of OS)
- Disk space: 10 GB recommended
- Windows 7