

Automatic Metadata Extraction

Usage scenarios – how EBU members use AME

This page contains an overview of some applications of Automated Metadata Extraction (AME), on which EBU members are working. The applications focus on different parts of media production and distribution workflows. Each application is introduced with a brief explanation, an overview of the technologies involved and a contact person for further information. These applications are driven by the media companies themselves and based on requirements from operational departments of the organisation.

This overview is open for contributions. If your company is currently working on an application of AME, it can be included in this list. Please contact the MIM-SCAIE chair Mike Matton <mike.matton@vrt.be> if you have an application which can be included in this list.

[Automated topic and keyword extraction for news \(VRT\)](#)

Scenario description: Two prototypes have been created to automatically enrich news articles and videos with metadata. The end goal is to improve search and retrieval, and to automatically identify links between different content items. Two prototype applications have been developed. The first prototype is a search engine with the enriched data of archived articles and video items. The second one is a tool for journalists. It automatically suggests tags for the articles they are writing. The intention is to integrate this in the Content Management System used by the journalists.

Technologies involved: Several enrichment algorithms are used for this purpose:

- IPTC-categorization: dividing articles or content into categories and topics
- A tagging algorithm that suggests the most important keywords
- Named Entity Recognition (NER)

Contact person: Marieke Lycke, VRT, marieke.lycke@vrt.be

[Automated subtitling workflows \(VRT\)](#)

Scenario description: During a 1,5 year long project called STON, the consortium partners have built a platform to automatize the subtitling workflow by means of speech and language technology.

Technologies involved:

- Speech detection: when are they speaking?
- Speaker clustering: who speaks when?
- Language detection: which language is being spoken?
- Speech recognizer: what is being said?
- Synchronization module for the alignment of script on video
- Post processing to stylize the subtitles according to the VRT standard

A cockpit is being developed to steer all components, to adjust the configuration to the needs of every program format and verify the results.

The focus within the STON-project lies on the automatized subtitling of documentaries and online news content, but the results of the project are also very valuable for other applications, such as the automatized annotation of audiovisual content.

consortium: Devoteam, KULeuven (Esat), Limecraft, UGent, PerVoice

contact person: Marieke Lycke, VRT, marieke.lycke@vrt.be

Automated tagging for web content (Yle)

Scenario description: The process of tagging web articles at Yle is computer aided: The journalist writes his/her article, the CMS sends it to an internal API (called Text Analysis API), which sends it to the SaaS company. The service returns tags (people, organizations, places, events and other subjects) which the journalist validates in CMS: he removes wrong and irrelevant tags and adds missing. The tags are used on Yle's web pages and in the news app for recommendations, navigation and personalization of services.

Several service providers have been tested "off line" and compared to the one used in production.

AME technologies involved: automated tagging on text

Contact person: Pia Virtanen, Yle Internet Media, pia.virtanen@yle.fi

Automated transcription of audiovisual content for recommendations (Yle)

Scenario description: Yle is currently piloting the use of automatic speech recognition for creating content based recommendations. The content item is first processed by a speech to text engine. The resulting text is then analyzed using statistical models to identify similarities across the content items. Finally, recommendations are created for each item.

Secondary use cases for transcriptions have been identified: SEO optimization by publishing the transcripts on the web, and automatic keyword identification.

AME technologies involved: ASR

Contact person: Kim Viljanen, Yle Areena, kim.viljanen@yle.fi

Automated tagging, content description and segmentation in video production (Yle)

Scenario description: Yle ran a PoC in 2016 where content description and segmentation was created by video AME service on current affairs productions, aiming to produce metadata currently entered manually. Visual concepts, objects and persons were recognized from the video, and footage was segmented based on visual analysis and semantic analysis on the subtitles.

AME technologies involved: visual content recognition, face recognition, text analysis

Contact person: Lauri Saarikoski, Yle Archives, lauri.saarikoski@yle.fi

ASR in automated subtitling for the deaf and hard of hearing (Yle)

Scenario description: Yle is obliged by law to subtitle nearly all domestic content into Finnish / Swedish, including live transmissions. Especially live content subtitling is very labour intensive without the help of speech recognition. Yle has therefore been testing different speech recognition models in their subtitling workflows.

Semi-automated subtitling by re-speaking instead of typing has been proven effective not only for live, but also offline content.

AME technologies involved: ASR

Contact person: Christoffer Forssell, Head of Yle Translation, christoffer.forssell@yle.fi

Automated tagging for image retrieval (NHK)

Scenario description: In order to make video footage available quickly and efficiently for the people who need and want to use them, annotating and indexing the material to make them searchable and retrievable is an important process. However, it is not feasible to perform

manual annotation if it is excessive in amount that requires a huge cost. Therefore, NHK has developed an automated tagging system for video footage. It has been validated on the "Great East Japan Earthquake" footage as a first trial target. Putting video footage on the storage will trigger a series of various annotation processes, such as shot detection, frame (image) extraction, low-level feature extraction, automatic tagging, generating indices. Finally, the system also provides shot-based fast search and retrieval function on the site.

AME technologies involved: Automated tagging, Automated low-level features extraction for similar image retrieval

Contact person: Masanori SANO (sano.m-fo@[nhk.or.jp](mailto:sano.m-fo@nhk.or.jp))

[ASR for generic subtitling \(NHK\)](#)

Scenario description: In order to broaden the range of programming available to people with different needs, including the elderly and hearing-impaired, subtitling (closed-caption) is one of the essential services. Generating a transcript is a labour intensive work however, especially for live broadcasting. NHK has developed an automatic speech recognition (ASR) system that produces real-time closed-caption for live news programmes with high accuracy and efficiency. It consists a hybrid ASR system that switches its speech input between the original program sound and speech rephrased by a "re-speaker". This solution is very cost-performant, and can be operated by only one or two staff members on the site.

As a further follow-up, NHK also began to develop algorithm for producing closed captions without manual effort, i.e. without staffs to correct recognition error. It compares news scripts prepared in advance with the recognition result, estimates the text corresponding to the speech and outputs the text as the closed captions.

AME technologies involved: Automatic Speech Recognition

Contact person: Shoei SATO (satou.s-gu@[nhk.or.jp](mailto:satou.s-gu@nhk.or.jp))