How to develop accurate and efficient methods for audiovisual content management

May 29 2020

Credit: Andrea Piacquadio from Pexels

Audiovisual media content is not only an essential tool for communication and entertainment, it is also seen as a useful source of modern history. To enable everyone to benefit from such informative

documents, it's crucial to translate moving images and sounds into words in an efficient and cost-effective way. Enter the EU-funded MeMAD project that is developing automatic language-based methods for managing, accessing and publishing pre-existing and originally produced digital content within the creative industries. Focusing on TV broadcasting and on-demand media services, the MeMAD project also aims to enhance digital storytelling.

Over two years into the project, MeMAD partners have developed a prototype platform to help audiovisual content professionals. They have also evaluated various aspects of the platform, as noted in a blog post on the project website. "There were four tracks of evaluation: video editing assistance, searching, intralingual subtitling with the help of automatic speech recognition (ASR) and interlingual subtitling with the help of machine translation (MT)."

The same blog post also states: "In all evaluations, participants filled out User Experience Questionnaire type forms (UEQ) after each task, adapted to focus on the task itself rather than the user interface. After each evaluation session there was a brief semi-structured interview. Additionally, in the video editing assistance and searching evaluations think-aloud data was collected." The blog post adds that ASR transcripts and MT were particularly found "to be useful in both video editing assistance and in archive searching, though there is still room for improvement."

In the same blog post, project partners emphasise that presentation and searchability of the metadata could also be improved. "It would be beneficial to be able to search for segments where a certain person speaks about a topic, combining face recognition and speech recognition data. Segment lengths also need to be examined, though in longer videos shorter segments may result in too many markers for a video editing tool to handle." The blog post continues: "Future evaluations will incorporate

face recognition and visual object detection. The same participants should be used in future evaluations, where possible, since they are already familiar with the platform and can compare results."

Novel tools

The MeMAD (Methods for Managing Audiovisual Data: Combining Automatic Efficiency with Human Accuracy) project uses machine learning and processing to enable easier retrieval of data from large volumes of content and from across several languages. It also aims to provide content creators with novel tools to better structure content and automate the delivery of content derivatives to various platforms like social media. Thanks to MT that is used for speech transcripts and subtitles, content will be available for new audiences in foreign languages and also be more accessible to people with hearing and/or vision impairment. One example for various use cases of MeMAD involves automated transcription, translation and subtitling. As a result, video editors editing interviews in a foreign language can work without the need for interpreters. In another use case, visually impaired consumers can follow current affairs shows with the help of auto-generated content descriptions that have been made available.

A periodic project report on CORDIS notes: "The key to the innovation is to provide Creative Industries with a common representation for the master data during the production processes, so that the current document-oriented editorial processes can be substituted with a more structured approach."

  More information: MeMAD project website: memad.eu/

Provided by CORDIS