Using Artificial Intelligence to Preserve Audiovisual Archives: New Horizons, More Questions

Jean Carrive

French National Audiovisual Institute (INA)

Bry-sur-Marne, France

jcarrive@ina.fr

ABSTRACT

France has a long tradition of preserving its archives as well as its cultural heritage, as demonstrated by the "legal deposit". Established in the Renaissance for printed documents, the legal deposit aims to allow the collection and consultation of various kinds of documents. INA, the French National Audiovisual Institute [1], is in charge of this task for France's radio and television, as well as French media on the web. INA's mission is to make the most of its collections: commercially by selling programs, and academically by making these collections available to researchers working on humanities and social sciences.

Since its creation in 1975, INA has constantly developed its tools and methodologies for describing and documenting its collections: databases, thesauri, lexicons, documentation software, indexing methods indexing, search engines, etc. Its Research and Innovation Department has for many years been interested in partnering with academic laboratories to explore the possibilities of automatic content analysis technologies. The emergence of AI-derived technologies is now making it possible to consider new uses of these collections, but also raises new questions.

INA has thus demonstrated that it now becomes possible to mass-treat large audiovisual corpora to identify various kind of information, thus facilitating indexing, documentation and search in order to provide better services to users. For researchers in humanities and social sciences working on these resources at Inathèque de France, these new means of analysis allow to conduct new types of Digital Humanities investigations, but also introduce new methodological challenges. For INA's archivists and librarians, AI's assistance facilitates the documentation process but also poses questions about the impact of these technologies on professional practices, as well as on the scalability of these technologies over time.

The presentation will address these questions, building on the research projects and experiments carried out at INA.
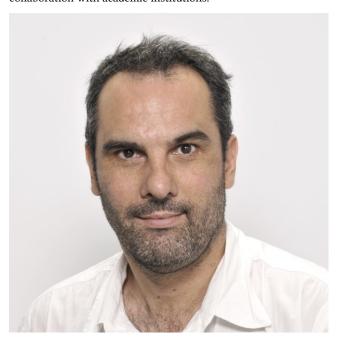
CCS Concepts/ACM Classifiers

• Information systems → Multimedia databases; Information systems → Multimedia and multimodal retrieval; Applied computing → Arts and humanities

BIOGRAPHY

Jean Carrive holds a Master Degree in Artificial Intelligence, Pattern Recognition and Application, and a PhD in Computer Sciences from University Pierre and Marie Curie, Paris 6 (now Sorbonne University). In his thesis "Classification of Audiovisual Sequences", he combined symbolic techniques from description logics on the one hand and constraint satisfaction techniques s on the other hand. He is now Deputy Head of Research and Innovation Department of the French National Institute of Audiovisual (INA) [1], a public institution dedicated to the preservation and the valorization of the French audiovisual heritage. He participated in or conducted several French or European projects in the area of automatic analysis of audiovisual contents: DiVAN, QUAERO, K-Space, InfoM@gic. He also supervised several PhD theses in this domain, in collaboration with academic institutions.

He now participates in the MeMAD H2020 Project (Methods for Managing Audiovisual Data) [2], which aims to develop automatic tools to facilitate access to audiovisual content, for example for people with disabilities. In the field of digital humanities, he is particularly interested in the application of audiovisual content analysis technologies for historical and heritage uses. In this regard, he is involved in the French

ANTRACT project (Transdisciplinary Analysis of French Newsreel) [3], which brings together historians and researchers in computer science with the aim of proposing a cross-approach on an emblematic audiovisual collection of the mid-twentieth century.

REFERENCES

[1] French National Audiovisual Institute, https://institut.ina.fr/en

[2] The MeMAD Project, http://memad.eu

[3] The ANTRACT Project, http://antract.hypotheses.org