AI-Driven Workflows for Unlocking Switzerland's Collective Memory: Distant listening of the RTS Archive

Giacomo Alliata, André Rattinger, Kirell Benzi, and Sarah Kenderdine

Laboratory of Experimental Museology, EPFL - first.last@epfl.ch

Keywords: Audiovisual Archives; Computational Augmentation; Natural Language Processing

Abstract

In the era of digital transformation, the vast digitization of audiovisual archives poses significant challenges for researchers in the arts and humanities (Fossati et al., 2012; Jaillant, 2022). This contribution presents a pioneering approach to address these challenges, utilizing cutting-edge AI technologies to design reproducible workflows for the efficient processing, analysis, and access to large audiovisual archives (Colavizza et al., 2021). Focusing on the case of the Radio Télévision Suisse (RTS) archive and its 200,000 hours of footage, essentially serving as Switzerland's collective memory, we demonstrate a comprehensive pipeline that integrates multiple algorithms and frameworks to facilitate the extraction of meaningful content.

The digitization of extensive audiovisual archives creates an overwhelming volume of data, making manual processing impractical. Our approach leverages the power of AI to operationalize these archives, unlocking new possibilities for researchers and the broader public to engage with the rich cultural and historical content within. The core components of our pipeline include speech-to-text conversion, speaker diarization, and named entities recognition (NER). These integrated algorithms work synergistically to transform raw audiovisual content into structured and annotated data.

The speech-to-text module plays a pivotal role in making spoken content accessible, allowing for the operationalization of speech within the archive. Thanks to the *Whisper* algorithm (Radford et al., 2023), we open audiovisual archives to the lens of NLP and "distant listening" enquiries (Clement, 2012, 2016). The speaker diarization component further enhances this process by segmenting the content into distinct "paragraphs", facilitating efficient navigation and analysis by extracting individual units of content. Incorporating NER adds a layer of richness to the data, connecting individuals, locations and events recorded in the archive to broader knowledge repositories. In particular, we have integrated the work previously described in Alliata et al. (2023) to connect the rich metadata extracted to open databases such as WikiData, expanding the full pipeline to the benefits of large knowledge bases (Rudnik et al., 2019).

The potential applications of our AI-driven workflows are diverse and impactful. One notable application is the creation of a dynamic "Map of Switzerland" visualizing the geographical distribution of voices recorded in the RTS archive, as shown in Figure 1. This serves as a novel exploration tool and contributes to a nuanced understanding of linguistic variations and cultural diversity over time, as well as geographical representations in the archive itself. Additionally, our workflows enable the implementation of Retrieval-Augmented Generation (RAG) pipelines (Lewis et al., 2020) to augment searchability within the archive leveraging Large Language Models and embedding models, enhancing the discoverability of specific topics, individuals, or historical events through more human-like queries.
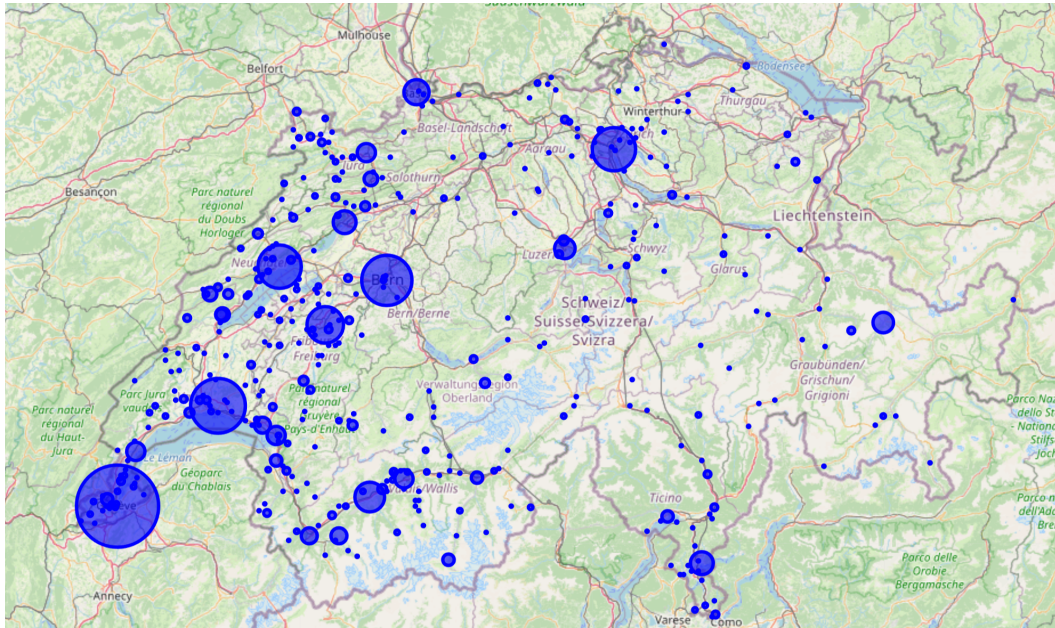
Figure 1: Geographical distribution of locations mentioned in the RTS archive (out of a sample of 77520 videos). The size of the circles is proportional to the number of occurrences, with Geneva having 4145 mentions.

This presentation emphasizes the technical, methodological and conceptual aspects of our AI-driven workflow, developed to operationalize large audiovisual archives through the dimension of spoken text. Such a "distant listening" approach can enrich these collections with a wealth of semantic information that opens up novel avenues of access and analysis. Furthermore, our case study on the RTS archive offers a compelling example of potential applications and thus contributes to the advancement of reproducible research practices in the arts and humanities.

Bibliography

Alliata, G., Yang, Y., & Kenderdine, S. (2023). Augmenting the metadata of audiovisual archives with NLP techniques: Challenges and solutions. *Digital Humanities 2023: Book of Abstracts.* https://doi.org/10.5281/zenodo.7961821

Clement, T. E. (2012). Distant listening: On data visualisations and noise in the digital humanities. *Digital Studies, 3.*

Clement, T. E. (2016). Towards a rationale of audio-text. *DHQ: Digital Humanities Quarterly, 10*(3).

Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2021). Archives and ai: An overview of current debates and future perspectives. *ACM Journal on Computing and Cultural Heritage (JOCCH), 15*(1), 1–15.

Fossati, G., et al. (2012). Found Footage Filmmaking, Film Archiving and new Participatory Platforms. *Found Footage. Cinema Exposed. Amsterdam: Amsterdam University Press/EYE Film Institute Netherlands.*

Jaillant, L. (2022). *Archives, access and artificial intelligence: Working with born-digital and digitized archival collections.* Bielefeld University Press.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. *International Conference on Machine Learning*, 28492–28518.

Rudnik, C., Ehrhart, T., Ferret, O., Teyssou, D., Troncy, R., & Tannier, X. (2019). Searching news articles using an event knowledge graph leveraged by wikidata. *Companion proceedings of the 2019 world wide web conference*, 1232–1239.