

Augmenting the Metadata of Audiovisual Archives with NLP Techniques: Challenges and Solutions

Giacomo Alliata, Yuchen Yang, and Sarah Kenderdine

Laboratory of Experimental Museology, EPFL

Keywords: Audiovisual Archives; Computational Augmentation; Natural Language Processing

Abstract

In the last decade, following the push of mass digitization, audiovisual archives have also been digitized, both by cultural institutions and broadcasting companies. The Radio Télévision Suisse (RTS) has for instance assembled more than 200,000 hours in its archive. However, as outlined by both archival scholars (Fossati et al., 2012) and the GLAM sector (Winesmith & Anderson, 2020), once digitized, these large collections require new frameworks to facilitate their access for the larger public and improve their civic value (Edmondson & Edmonson, 2004). Actually working with such large datasets however presents several challenges, that this research sets to highlight and then address. In this paper, we tackle these challenges from the perspective of experimental museology (Kenderdine, 2021), with the goal of augmenting the metadata of the RTS Archives for supporting innovative modes of access through Natural Language Processing (NLP) techniques.

The main difficulty that arises is the potential lack of homogeneity in the archival process. This can result in overwhelming amounts of different tags and categories, unusable as is to actually cluster the items in the collection in a meaningful way. For instance, in the RTS Archives, we have 6528 topic tags, 3273 geographical tags and 8954 entity tags across 522103 videos. Figure 1 demonstrates that the distribution of the number of tags per video is heavily right-skewed, with 75% of the videos having a maximum of 4 tags.

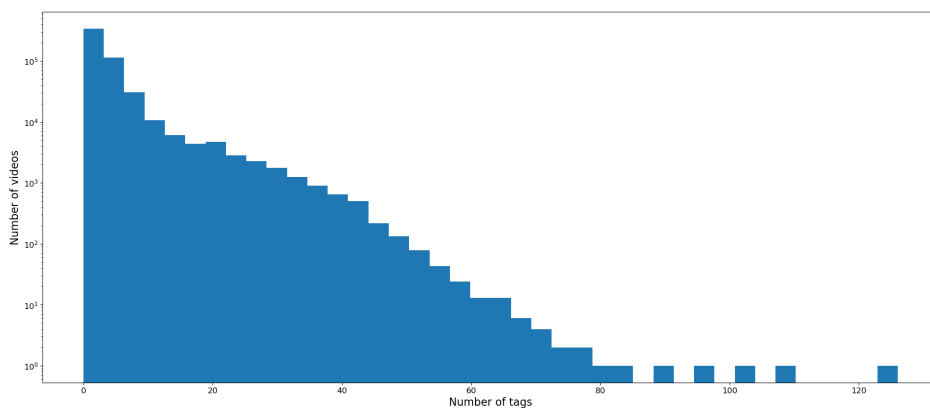


Figure 1: Distribution of the number of tags used per video in the RTS Archives (log scale).

and linked to a Wikidata entity, the RTS metadata automatically inherits the semantic and logical information from the Wiki knowledge graph. The audiovisual archive can thus be navigated through the Wikidata database. Properties of these entities can be queried in a simple manner, obtaining geographical coordinates for the location or citizenship of the persons for instance. The drawback of having a manual and arbitrary tagging system (often found missing tags, inconsistency in multi-lingual tags, etc.) is also solved by providing a holistic and connected higher structure to the tags.

In conclusion, this research shows how the metadata of an audiovisual archive tagged in an unconstrained manner can be augmented by linking it to a well-structured and large collection: the Wikidata database. Future work will explore how tags can be automatically extracted from the videos, transcribing the audio for instance, and thus potentially resulting in a much more granular tagging, on the scale of short clips rather than entire videos.

Bibliography

- Barba, E., Procopio, L., & Navigli, R. (2022). ExtEnD: Extractive entity disambiguation. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.
- Edmondson, R., & Edmonson, R. (2004). *Audiovisual Archiving: Philosophy and Principles*. UNESCO Paris.
- 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.
- Kenderdine, S. (2021). Experimental Museology: Immersive Visualisation and Cultural (Big) Data. *Experimental Museology*.
- Rodighiero, D., Derry, L., Duhaime, D., Kruguer, J., Mueller, M. C., Pietsch, C., Schnapp, J. T., & Steward, J. (2022). Surprise machines: Revealing harvard art museums' image collection. *Information Design Journal*, 27(1), 21–34.
- Winesmith, K., & Anderson, S. (2020). *The Digital Future of Museums: Conversations and Provocations*. Routledge.