PROPOSAL FOR AN APPROACH TO PROCESSING MEDIA OBJECTS USING AI AND METADATA: A MACHINE LEARNING-BASED APPROACH

Khalifa Sylla[1], Mama Amar[1] and Samuel Ouya[2]
[1]Cheikh Hamidou Kane Digital University of Senegal Un-CHK (ex-UVS),
Cité du Savoir, Diamniadio, 15126 Dakar-Fann, Senegal
[2]LITA Laboratory, Polytechnic High School, Cheikh Anta Diop University of Senegal, Senegal

ABSTRACT

The paper must have an abstract. The abstract should be self-contained and understandable by a general reader outside the context of the paper. The use of artificial intelligence (AI) has revolutionized the organization and management of data in particular educational content on digital university platforms.
The combination of AI and metadata can facilitate the management and access to educational resources and allow personalized learning experiences tailored to learners' needs. Based on the possibilities offered by AI and metadata, we propose, in this paper, a model for processing media objects using AI and metadata. Our approach is based on the use of machine learning to enrich media objects with original and extracted metadata and efficiently process media objects (images, videos, etc.) using both raw media data and associated metadata.
In this study, we base use the pedagogical resources of the Cheikh Hamidou Kane Digital University's training courses.
The proposed model enables us to extract contextual information, classify metadata and categorize multimedia objects for future use.
The application of our model makes it possible to classify educational data by category or activity (consultation, TD, TP, project) according to pedagogical objectives; to automatically add metadata to a media object using the model; to manually annotate a media object by pedagogical actors; and finally, to create a search engine based on metadata.

KEYWORDS

Metadata, IA, Machine Learning, Multimedia, Annotations, Process

1. INTRODUCTION

The development of computing power and storage capacity encourages the multiplication of the production of multimedia documents. Thus, we are observing more and more creation of multimedia content in all sectors and in the field of education. Today, several digital universities are created (MOOC platforms, Digital Universities, e-learning systems, etc.). Since media objects are not structured, their exploitation generates a certain complexity.

In addition, we note that metadata offers additional information which can, through their exploitation, enrich multimedia objects and help in their exploitation. However, extracting and leveraging this metadata can be costly and time-consuming for organizations. Indeed, faced with the exponential growth of multimedia production, the various players in this sector are faced with numerous problems regarding the storage, annotation and extraction of multimedia objects. Thus, the international standards organization (ISO) and other multimedia communities have developed standards such as MPEG-7 (Sikora, 2001), Dublin Core for the description of the content of multimedia objects. However, these standards are not widely used because it is difficult to manually annotate multimedia objects.

On the other hand, artificial intelligence has seen significant development over the past decade. It now offers advanced capabilities for exploiting large volumes of data. These capabilities enable analysis and extraction of relevant information at a scale and speed previously impossible. As a result, the use of machine learning will make it even easier to automate the production of metadata to help find, filter and organize information, particularly multimedia type information.

In this article, we explore the possibilities offered by artificial intelligence and metadata associated with multimedia objects for optimal management of media objects. This will include developing machine learning models to extract contextual information, classify metadata and classify media objects for future exploitation. In this study, we base ourselves on the training content at the Cheikh Hamidou Kane Digital University (formally Virtual university of Senegal) (Ouya et al., 2015).

This paper will be organized as follows: we will start with the state of the art. Next, we will present our approach. Additionally, we will outline our proposed machine learning models. We will continue with the presentation of the implementation concept, then we will present the results of our experiments. Finally, we will end with the conclusion and perspectives.

2. STATE OF THE ART ON MEDIA OBJECTS MANAGEMENT MODELS

2.1 Multimedia Metadata

2.1.1 Metadata

A metadata is a data or information used to describe other data. There are several standards used to describe or classify media data using metadata. In the context of a library, metadata traditionally refers to elements such as bibliographic information about the collection's holdings and physical items; however, the current definition of metadata has evolved to include information about electronic or digital data, rather than primarily describing physical data elements.

Metadata can be classified by type to facilitate their management and organization. Thus, in our work, we have chosen the following types of metadata that are adapted to our use case, which will be presented in the following sections. However, our model allows for the definition of other types of metadata as well. We have identified below metadata types (Amar, 2012):

- ***Descriptive metadata*** include the metadata used to describe media objects (MOs) (Kogalovsky, 2013) and allow for the search and identification of an MO. Examples: title, author, location.
- ***Technical metadata*** describe the technical characteristics of media objects that will be used to classify the MOs. Examples: file format, size, dimensions.
- ***Administrative metadata*** represent the administrative information of the MOs and can be used in the management and archiving of an MO. Examples: archive date, rights information, archive profile, version.

2.1.2 Multimedia Metadata Standards

Metadata standards allow for the formalization of a metadata model that describes the semantics and value lists of the description elements as well as the links between the description elements. In the literature, we find several metadata standards, including:

Dublin Core (Underwood, 2020), which was developed to describe electronic text documents but has been extended to cover media objects.

It proposes a generic metadata schema that allows for the description of digital or physical resources based on a set of elements.

MPEG-7 (Multimedia Content Description Interface) (Burnett et al., 2003)(Morgan, 2012) is an international standard for describing multimedia objects recognized by the ISO/IEC consortium and developed by MPEG (Moving Picture Experts Group). The standard focuses on the semantically rich description of multimedia content. MPEG-7 allows the localization, filtering, management, and processing of desired multimedia files by tagging them with information about their content and origin.

MPEG-21 (Burnett et al., 2003) was developed to ensure the interoperability of multimedia objects. It is based on two essential concepts: the Digital Item, which represents the basic component of the standard, and the concept of Users interacting with the Digital Item. The Digital Item is a structured object with a standardized representation that includes multimedia content, associated metadata, and a set of syntax elements that describe the relationships between the media objects and their metadata. The user concept represents any entity interacting with the MPEG-21 environment or the Digital Item.

SMPTE Standard (Morgan, 2012) is a metadata dictionary standardized by the Society of Motion Picture and Television Engineers (SMPTE). SMPTE offers a list of structured metadata elements. A metadata dictionary is used to define the keys, the length of the values, and the semantics of these elements. The current structure defines classes of elements as follows: identification, administration, interpretation, parametric, process, relational, and spatio-temporal.

2.2 Metadata and ML for Multimedia Objects Processing

In this section, we will analyze a set of articles that use machine learning and metadata, selected based on their publication date (recency), the AI techniques used, and the concepts employed.

Studies on the use of metadata and AI for the management of media objects discuss a method for automatic classification of MOOC video topics using transcription features and convolutional neural networks (CNNs) (Chatbri et al., 2017). The method involves generating video transcripts using a speech recognition tool and converting the transcripts into images using a statistical co-occurrence transformation. The CNN (Sun et al., 2020) is then used to produce video category labels for the transcript image input. The method is evaluated using a Khan Academy dataset, which contains 2,545 videos labeled with one or two of 13 categories. The results show that the proposed method outperforms a baseline algorithm that uses superficial features, demonstrating its effectiveness in automatic video classification.

This method does not use video metadata, which could provide more information to facilitate the classification of course videos.

S. Bharitkar presents a method for classifying multimedia content using synthetic metadata and machine learning (Bharitkar, 2019). Synthetic metadata is used to train a neural network to classify multimedia content, achieving high accuracy with low latency. Synthetic metadata improves classification accuracy by providing additional features that complement the original content.

3. PROPOSAL OF A MODEL

The model we propose is based on the extraction, annotation, storage, and archiving of media objects. This approach aims to enable optimal archiving and utilization of media objects.

The model allows for the extraction and classification of metadata, annotation of media objects, utilization, and archiving after a certain period. This approach leverages the metadata associated with these objects to facilitate the organization, storage, archiving, and retrieval of these objects.

We also present the application that served as the framework for our approach. This approach is based on existing tools and provides solutions to address the problems or shortcomings of these tools.

3.1 Presentation of Our Approach

Our approach is based on a process that covers all stages of the life cycle of a media object (MO), from creation to archiving, and on the metadata associated with these objects. Throughout the MO's life cycle, additional metadata can be created and linked to the object. This metadata provides more information about the MO, thereby facilitating search, management, and archiving. This approach, while presenting specific features compared to existing approaches, incorporates some of their basic functionalities.

3.2 Use Case Definition

Our model is applied to course videos used in the teachings at UN-CHK. Our objective is to enable better utilization of resources (videos, syllabi, courses, supplementary documentation) by students and teachers by allowing them to:

- Classify by category or activity (consultation, tutorial, lab, project) based on educational objectives.
- Automatically add metadata to a media object by the model.

- Manually annotate a media object by the educational stakeholders.
- Use a search engine based on metadata.

3.3 Description of our Approach

The model we propose is based on a process composed of the following phases: Collection of media objects, media preprocessing, processing and extraction of metadata, annotation of media objects, and classification of media objects for future use.

- **Collection and Storage of Media Objects:** This phase involves gathering and importing educational data from various sources to apply our approach for metadata extraction. Once the media objects have been collected, they are stored in a database or file system associated with metadata for quick search and access.
- **Media Preprocessing:** This phase involves standardizing and normalizing the media objects to ensure they are understandable by our system.
- **Processing and Extraction of Metadata:** This phase involves extracting information that allows the description and classification of media data. During this phase, only the metadata that comes with the media object at its creation is extracted. The extracted metadata will be classified according to predefined categories.
- **Annotation of Media Objects:** This phase involves adding new metadata to be used effectively. There are two methods of annotation: the first involves manually adding metadata to the media object, and the second uses a machine learning algorithm to extract metadata from the media object and classify it according to predefined categories.

4. DESIGN AND IMPLEMENTATION

4.1 Description of the Model

In this section, we will present the approach we propose. We will outline the design and development of our model based on metadata and machine learning.

The approach we propose uses machine learning to efficiently process media objects (images, videos, etc.) by utilizing both raw media data and associated metadata. The approach is based on the following phases:

Collection and Storage of Media Objects: This phase categorizes media objects based on their content. In this step, we will use media objects created in the context of educational activities at UN-CHK (Sylla et al., 2020).

Preprocessing of Media Objects: This phase analyzes and extracts relevant information (metadata) from the content and characteristics of the media objects, storing it in a database for future use with the media objects.

Processing and Extraction of Metadata: Media objects are processed to extract metadata, which is then stored and associated with the media objects for future use. Metadata is subsequently processed (normalizing values, managing missing data, and transforming categorical data into numerical format) to organize and clean it.

Classification of Media Objects: A classification algorithm is used to classify media objects according to their types and categories.

4.2 Architecture of the Model

Our architecture is largely inspired by the machine learning lifecycle diagram proposed by Amazon Web Services (AWS) (Wittig et al., 2023). Figure 1 below shows the different components of this lifecycle.

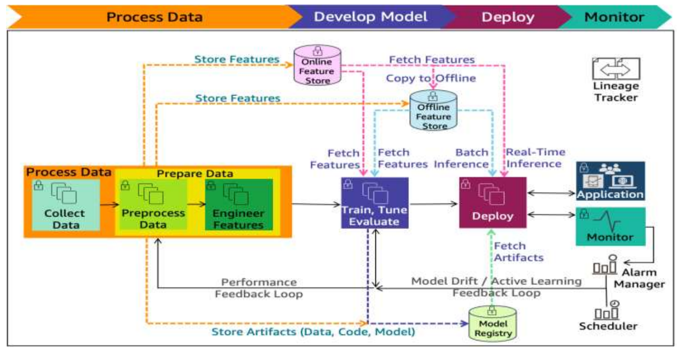

Figure 1. Different components of the lifecycle of ML models deployment

Thus, the architecture of the model we propose is structured around the following components:

- Data acquisition and processing: Data acquisition and processing involves collecting and preparing the data.
- Data preparation includes data preprocessing and feature engineering.
- Model development: Model development includes creating, training, tuning, and evaluating the model.
- Model deployment: After training, tuning, and evaluating our model, it will be deployed in production to be used with our production data.
- Model monitoring: The model monitoring system captures data, compares it with the model's training data, sets rules to detect issues, and sends alerts.

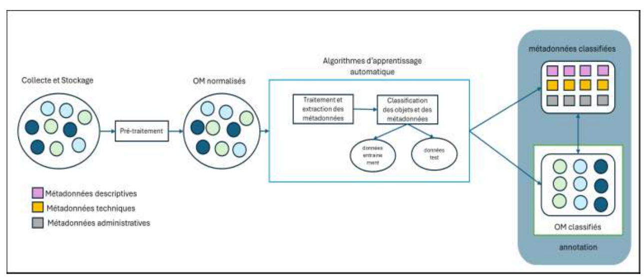

Figure 2. Architecture of the models of our solution

4.2.1 Data Collection and Preparation

The data to be used consists of a set of MOs (Media Objects) from online courses at UNCHK, including the MOs and their associated metadata. The data is cleaned to remove unnecessary or redundant information, handle missing data, and normalize the data if necessary.

4.2.2 Introduction to Algorithms in Our Model

In this stage of our process, we will perform data splitting. We will divide the dataset into two parts: training data and test data to evaluate the performance of the algorithms composed of the metadata extraction and classification algorithm, and the media object classification algorithm.

a) Metadata Processing Algorithm

For metadata processing, we have chosen to use the Convolutional Neural Network (CNN) algorithm. CNNs are commonly used to extract significant features from images and videos, making them suitable for extracting visual metadata such as objects, faces, and locations. CNNs can also be used for image classification tasks based on visual metadata.

The algorithm begins by examining each file in the training data directory to extract metadata. Once extracted, the algorithm classifies them into the predefined categories: Descriptive, Technical, and

Administrative. After classification, the metadata is stored in a DynamoDB database along with their links to the multimedia objects, ready for analysis or export in formats such as CSV.

While this algorithm focuses on metadata extraction and classification, it will also be integrated into the Machine Learning pipeline with the MO processing algorithm outlined in our approach for tasks such as content recommendation and multimedia data analysis.

The algorithm is designed to be adaptable based on specific user needs. The metadata extraction and classification functions can be customized by adding new features or adjusting classification criteria.

```
# Extraction des Métadonnées avec Apache Tika :
from tika import parser
import os
import pandas as pd
import csv


# Main function
def main():
    directory_path = './medias'  # directory of the media objects
    output_csv_filename = 'metadata_output.csv'
    metadata_list = []

    # Iterate through files in the directory
    for filename in os.listdir(directory_path):
        file_path = os.path.join(directory_path, filename)

        # Ensure it's a file (not a subdirectory)
        if os.path.isfile(file_path):
            metadata = extract_metadata(file_path)
            descriptive, technical, administrative =
classify_metadata(metadata)
            metadata_entry = {
                'File': filename,
                'Descriptive Metadata': descriptive,
                'Technical Metadata': technical,
                'Administrative Metadata': administrative
            }
            metadata_list.append(metadata_entry)

    # Save metadata to CSV
    save_metadata_to_csv(metadata_list, output_csv_filename)
```

Figure 3. Metadata Extraction Algorithm

b) Media Object Processing Algorithm

For processing media objects, the Support Vector Machine (SVM) algorithm (Dai, 2018) has been chosen for the following reasons:

- SVMs are effective for binary or multi-class classification.
- They can be used to classify media objects based on their extracted metadata.
- SVMs are robust and perform well, especially with medium to large-sized datasets.

The multimedia data is preprocessed to be compatible with the SVM algorithm. This includes resizing, normalization, and extracting relevant features from the media.

The SVM model is trained on a labeled training dataset. Features extracted from the media objects and their associated metadata are used as input vectors, while class labels (types of media objects) are used as outputs.

Once trained, the SVM model is used to classify new media objects based on their predicted types. Classification is based on both media features and associated metadata.

The SVM model can be used to recommend relevant content to users based on their preferences and analysis of similar media objects.

c) Model Training and Validation

Our model is trained on a training dataset to adjust model parameters using extracted metadata as input and class labels as output. We evaluate the performance of our model using the accuracy metric to assess the quality of the classification.

5. RESULTS OF OUR SOLUTION

The results of our model can be presented according to the following components: Classification of Media Objects by Category and/or Activity Based on Educational Objectives.

Objective: Classify media objects (images, videos, audios) based on educational objectives, such as consultation, tutorials (TD), practical work (TP), and projects.

Results:

- Media objects are automatically classified into one of the educational categories (consultation, TD, TP, project).
- Improved organization of educational resources and easier access for students and teachers.

Addition of Metadata to a Media Object by the Machine Learning Model:

Objective: Enrich media objects with metadata automatically generated by the machine learning model.

Results:

- Media objects have richer and more accurate metadata, improving their descriptions and facilitating their search and use.
- Reduced time and effort required for manual annotation of media objects.

6. CONCLUSION

In this paper, we proposed an ML based solution that used metadata for efficient processing of media objects. The metadata makes it possible to enrich media objects and thus provide better processing of these unstructured data. With machine learning, our solution offers the possibility not only to process and classify metadata, but also to generate them automatically and use them to enrich media objects. In this work, we first defined a process composed of several steps which allow us to collect the OM, to pre-process the OM, to extract and process the metadata and to process and classify the OM using the metadata and machine learning.

To validate our approach, we experiment our solution using multimedia objects from eLearning files from UNCHK. The results of our approach showed the enrichment of media objects from original and extracted metadata from the media objects.

Our approach also improves the organization of educational resources and easier access to them for students and teachers.

As future work, we want to add more functionalities to our solution by allowing users to manually annotate multimedia objects. We also want to include a multimedia search engine that uses the metadata to our solution.

The approach that we proposed is generic and the solution can be applied to other use cases and domains. Thus, we want to use our solution for another use case. Ex: We want to use our approach to optimize the search for scientific articles in scientific article journals.

Furthermore, we would like to include a recommender system module in our solution to recommend e-learning videos to the students based on the related metadata.

We also want to experiment our solution with other machine learning models and perform comparisons to validate the performance of our model with different ML algorithms.

REFERENCES

Amar, M. (2012). *Métadonnées et processus pour l'archivage de données médiatiques* (Master's thesis). Université du Québec à Montréal, Montréal, Canada.

Burnett, I., Van de Walle, R., Hill, K., Bormans, J., & Pereira, F. (2003). MPEG-21: goals and achievements. *IEEE Multimedia*, 10, 60-70. https://doi.ieeecomputersociety.org/10.1109/MMUL.2003.1237551

Chatbri, H., McGuinness, K., Little, S., Zhou, J., Kameyama, K., Kwan, P., & O'Connor, N. E. (2017). Automatic MOOC video classification using transcript features and convolutional neural networks. In *Proceedings of the 2017 ACM Workshop on Multimedia-based Educational and Knowledge Technologies for Personalized and Social Online Training* (pp. 21-26). https://doi.org/10.1145/3132390.3132393

Dai, H. (2018). Research on SVM improved algorithm for large data classification. In *Proceedings of the IEEE 3rd International Conference on Big Data Analysis (ICBDA)*. https://doi.org/10.1109/ICBDA.2018.8367673

Kogalovsky, M. R. (2013). Metadata in computer systems. *Programming and Computer Software*, 39, 182-193. http://dx.doi.org/10.1134/S0361768813040038

Morgan, O. (2012). Structure and organization of metadata and the SMPTE metadata dictionary. In *File Interchange Handbook* (pp. 31-60). Routledge.

Ouya, S., Sylla, K., Sy, B., & Barry, M. (2015). Impact of integrating WebRTC in universities' e-learning platforms. In *Proceedings of the 2015 5th World Congress on Information and Communication Technologies (WICT)*. https://doi.org/10.1109/WICT.2015.7489664

Pereira, F. (2005). MPEG-21 standard: defining an open multimedia framework. In *Proceedings of the 47th International Symposium ELMAR*. Zadar, Croatia, pp. 5-8. https://doi.org/10.1109/ELMAR.2005.193627

Sikora, T. (2001). The MPEG-7 visual standard for content description-an overview. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 696-702. https://doi.org/10.1109/76.927422

Sun, Y., Xue, B., Zhang, M., & Yen, G. G. (2020). Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE Transactions on Cybernetics*, 50(9), 3840-3854. https://doi.org/10.1109/TCYB.2020.2983860

Sylla, K., Sy, B., Ouya, S., & Barry, M. (2020). Towards the use of a contact center for the socialization and capacity reinforcement of learners of African digital universities. In *Proceedings of the 22nd International Conference on Advanced Communication Technology (ICACT)*. https://doi.org/10.23919/ICACT48636.2020.9061548

Underwood, W. (2020). Automatic extraction of Dublin Core metadata from presidential e-records. In *Proceedings of the 2020 IEEE International Conference on Big Data (Big Data)*, pp. 1931-1938. https://doi.org/10.1109/BigData50022.2020.9377943

Wittig, A., & Wittig, M. (2023). *Amazon Web Services in Action: An in-depth guide to AWS*. Simon and Schuster.