Abstract—Being faced with digital libraries containing images and video content, means are required that characterise the content for efficient access. In addition, the confluence of media content which is distributed over a number of content providers requires a common and standardised way for searching the content. The usual solution consists in the employment of metadata which describes media content. Among others, issues that arise concern the kind of metadata to be used, how it is to be represented, and how it is to be integrated into the architecture of a portal for accessing the distributed digital archives. The methods presented in this paper have been implemented in the context of the project Gateway to Archives of Media Art (GAMA for short). The objective of this project is to establish a portal for online access to some of the most important digital archives and libraries on media art in Europe.

I. INTRODUCTION

Through the GAMA portal media art content is made accessible for curators, artists, academics, researchers, and mediators as well as for the interested public. It is mainly video art that is to be made accessible within this project. Thereby, metadata plays an important role in advanced search that includes Query by Example (QbE hereafter) and further subsystems for the analysis of the content of videos and still images. Moreover, interfaces enable to browse through the content via criteria such as artists, genre, or production year.

While the provision of full, high-quality content is up to the content providers, the idea behind the GAMA system is to provide metadata from the local databases of the different content providers,

1) to collect metadata from the local databases of the different content providers,
2) to store the metadata in a centralised repository,
3) to perform content-based analysis of media content,
4) to produce audiovisual descriptions, and
5) to provide centralised searching and browsing functionalities, making use of all connected archives.

There are several similar portals which have been devised for different purposes in the past. Their evaluation by the GAMA content partners however show that emphasis can be put on different aspects and such systems are normally tailored to specific needs of their users. For the GAMA project, of most interest are: the UBU portal\(^1\) that offers well structured material and multilingual access, however the metadata is very limited; the VideoArtWorld Homepage\(^2\) allows the content to be searched by categories and provides a transparent structure; but here again metadata about the content is very limited and the search functionality is rather simple; the New Media Encyclopedia\(^3\) is distinguished by a nice glossary and well structured bibliography, but suffers from a complex menu structure and mixes categories such as general expressions and names of artists. Other sites like the homepage of the Daniel Langlois Fondation\(^4\) are also of interest in that they also provide multimedia content, but metadata information of interest for content users of media art require other modes to search the contents. A main objection that is to be raised is that the visual access is very limited in current systems.

II. MEDIA CONTENT AND METADATA

Media content is enriched with metadata mainly in two ways. On the one hand, metadata is directly imported from the databases of the content providers; it is then the challenge to integrate data from different sources with heterogeneous data models and to ensure interoperability. On the other hand, content-based metadata is extracted from the raw media content, such as descriptions of audiovisual characteristics of media content; this is typically not available from the content providers and extracted by the central content-based indexing service. Figure 1 displays relevant parts of the GAMA architecture.

III. GAMA SYSTEM ARCHITECTURE

Aim of the GAMA system architecture is to offer the content providers a clear and simple to use infrastructure.

---

\(^1\)See http://ubu.clc.wvu.edu

\(^2\)See http://www.videoartworld.com

\(^3\)See http://www.newmedia-art.org

\(^4\)See http://www.fondation-langlois.org
Thereby, the required infrastructure of the different archives should be minimised. In the simplest case, it would not even be necessary for content providers to have a permanent internet connection or to give direct access to their media or database servers. In this case it would be possible to upload media and metadata (e.g. database dumps) to the GAMA servers. This is important inasmuch part of the content providers do not allow their servers to be directly accessed for legal reasons. Furthermore, no software component needs to run at the side of the content providers; rather, the GAMA system architecture is organised by services which run at distributed locations. Relevant in the context of this paper are services dealing with metadata; these are the content-based indexing service and the central metadata repository.

The central metadata repository is the central storage component for all metadata in the GAMA system. It provides a SOAP interface for data ingest and for querying the repository. Metadata available from the archives is either automatically downloaded from the servers at the archives or uploaded by content providers and mapped by database adaptors.

The content-based indexing service analyses raw media and generates content-based descriptions (see Section V). Additionally there are two QbE subsystems based on the extracted features (see Section VI). Media is either automatically downloaded from the servers of the archives or uploaded by content providers. Metadata generated by the content-based indexing service (e.g. audiovisual descriptions) is exported to the central metadata repository through the SOAP interface for data ingest. Part of the output is file-based, e.g. thumbnails of key frames that are extracted per video shot. File-based output is written directly to the file storage.

Both metadata generated by the content-based indexing service and metadata imported from the archives is accessed by the GAMA portal through the query interface of the central metadata repository (see Section VI). File-based outputs of content-based analysis (e.g. thumbnails of key frames) are accessible for use within the portal through a web server via HTTP.

**IV. Metadata Import from Local Databases of the Content Providers**

Metadata from local databases of the content providers is imported into the central repository through so-called database adaptors (see Figure 1). The data model of the central repository is based on RDF\(^5\), which is a flexible solution with regard to heterogeneous data models of the archives. Database adaptor implementations are content provider specific and map data from local databases to RDF/XML according to the GAMA RDF schema, and transfer it to the central GAMA metadata repository.

---

\(^5\)Resource Description Framework, see http://www.w3.org/RDF
Metadata from the databases of the content providers includes information on artwork (e.g. title, date of creation), manifestations of artwork (e.g. format, length), persons such as artists or curators (e.g. name, date of birth), archives (e.g. name, homepage), and similar and relations thereof, such as “an artwork is provided by an archive” or “a person (artist in that case) is author of an artwork”.

V. CONTENT-BASED ANALYSIS AND EXTRACTION OF AUDIOVISUAL DESCRIPTIONS

Content-based metadata is extracted by the content-based indexing service which has a module-based structure. Several modules extract content-based features, e.g. Automatic Speech Recognition (ASR), still image and video Optical Character Recognition (OCR), face recognition, shot boundary detection, as well as extraction of a set of MPEG-7 visual, audio, and video-specific descriptors. Part of the extracted content-based metadata is textual data that directly allows for text-based querying, e.g. results of ASR and OCR. Other metadata is utilised within the two QbE subsystems for image and video content. Additionally, key frames are extracted from a shot and used as elements to visually enhance the GAMA user interface.

Feature extraction modules applied within the content-based analysis of the GAMA system are described in the following sections.

A. Shot boundary detection and key frame extraction

The shot boundary detection and key frame extraction module based on the approach described in [1] extracts shot boundaries within input video content and thereby extracts representative frames per shot. Results have the form of start and end frame number per shot and associated images (key frames). Key frames are extracted with three different resolutions:

- original resolution of input video
- large image (fixed-size for use in the portal)
- thumbnail (fixed-size for use in the portal)

Extracted shots are the basis for video-based QbE within the GAMA system. Video features are extracted and matched on a shot basis. Queries for video QbE are shots for which associated key frames are displayed in the GAMA portal and linked with QbE functionality. Extracted key frames in original resolution are fed into the analysis process for feature extraction and used as a part of video QbE.

Key frames can also be used in various ways as visual elements within the GAMA portal, e.g. in result lists or in a detailed view of a video to give an overview of the temporal structure.

B. MPEG-7 audiovisual descriptors

The MPEG-7 (“Multimedia content description interface”) standard has been selected for describing the audiovisual content. The descriptions are the result of the analysis of both domains the visual and the audio.

For MPEG-7 based visual [2] indexing a subset of MPEG-7 visual descriptors is chosen. MPEG-7 visual descriptors can further be divided into descriptors applicable for still images/video frames and descriptors that are extracted on the basis of video segments, respectively shots. For still images/video frames the following MPEG-7 visual descriptors are extracted:

- Color Layout Descriptor
- Dominant Color Descriptor
- Scalable Color Descriptor
- Color Structure Descriptor
- Edge Histogram Descriptor

These descriptors are utilised for both still-image matching and video matching [3] within the respective QbE subsystems (see Figure 2 with an example utilising images downloaded from Flickr⁶). Within the video QbE subsystem these descriptors are extracted from key frames extracted by the shot boundary detection module (see last paragraph).

As the goal of the MPEG-7-based QbE subsystem is to provide the best matches to the query object, currently some user tests are executed, which aim at selecting the best combination of the above mentioned visual descriptors. The user tests are tailored to the media art content and consumers. The tests will enable the construction of the optimal model for combining the MPEG-7 visual descriptors.

Additionally, for video QbE there are video-specific visual and audio descriptors extracted on a shot basis. Two video-specific visual descriptors are extracted per video shot:

- Motion Activity Descriptor
- Camera Motion Descriptor

Finally two audio descriptors, [4], are extracted per video shot. These are:

- Audio Spectrum Centroid Descriptor
- Audio Power Descriptor

MPEG-7 audiovisual descriptions are utilised within the QbE subsystems for still images and videos. For comparison of descriptors the distance measures proposed by the MPEG-7 standard [3] are applied.

C. TZI PictureFinder

TZI PictureFinder [5] is an extremely fast matching solution for image-to-image (or frame-to-frame) matching based on the distribution of visual features such as color and texture. It is especially optimised for matching within large datasets. In the context of the GAMA system TZI PictureFinder is used as a pre-filtering solution within both QbE subsystems to considerably improve the performance of the generation of pre-calculated result lists that are then exported to the central RDF repository (see Section VI).

D. Optical Character Recognition (OCR)

The GAMA OCR module extracts text from video frames and still images. For input videos OCR is applied on every n-th video frame. OCR results are further filtered to avoid

⁶See http://www.flickr.com/
misdetections. In the GAMA system Tesseract OCR\(^7\), a freely available open source optical character recognition engine, is deployed. Results are words occurring in still images or video frames and time of occurrence (on a shot basis) in case of videos. The OCR module produces textual output that directly enables text-based queries.

**E. Automatic Speech Recognition (ASR)**

The GAMA ASR module extracts spoken text from audio tracks of input videos. The Microsoft Speech Application Programming Interface (SAPI)\(^8\) is used. Results are spoken words and time of occurrence (on a shot basis). The ASR module produces textual output that also enables text-based queries.

**F. Face Recognition (FR)**

The GAMA FR module detects faces occurring in input videos and recognises faces which appear repeatedly. The CMU Face Detector, \([6]\), is employed for detecting faces whilst the CSU Face Identification Evaluation System \([7]\) provides standard face recognition algorithms and statistical methods for comparing face recognition methods. Results are identifiers of actors and the time of occurrence in a video. This enables the search for occurrences of the same person (actor) in other videos or other parts of a video for a given actor.

**G. Sound Event Detection and Music/Speaker Segmentation**

The sound event detection module based on \([8]\) is a generic component that can be trained to detect certain predefined audio events based on low-level audio features and their classification through the application of Support Vector Machines. In the context of GAMA it will be trained for a number of events that are currently discussed with content partners. It enables the search for video segments where certain audio events occur.

In the context of the SVP project\(^9\), \([9]\), it has been successfully applied to identify video segments with spoken words or background music. These models will be applied within GAMA to realise a music/speaker audio segmentation. The latter looks for segments with spoken text and/or background music in the audio tracks of input videos. It will be used as a feature within video QbE.

**VI. Querying the RDF Repository and QbE Subsystems**

The common query interface is the SPARQL\(^10\) query interface of the central RDF repository. It is employed for metadata

---

\(^7\)See \[http://code.google.com/p/tesseract-ocr\]

\(^8\)See \[http://www.microsoft.com/speech/speech2007/default.mspx\]

\(^9\)See \[http://www.tzi.de/svp\]

\(^10\)SPARQL Protocol and RDF Query Language, see \[http://www.w3.org/TR/rdf-sparql-query\]
from the databases of the archives which are collected by
database adaptors; additionally, it is employed for metadata
which have been automatically extracted from content-based
indexing modules; this concerns especially QbE for images
and videos. For the sake of efficiency (regarding response time
and expected traffic on the GAMA portal) QbE operations
are not performed online. Instead, ordered lists of a fixed
number of results are pre-computed and stored within the RDF
repository for every potential query (every shot of a video
indexed by the GAMA content-based indexing service) and
updated on a regular basis. This is applied for both images as
well as videos.

An important advantage of this approach is that QbE results
can directly be aligned with other search criteria through the
same query interface (e.g. search for artwork of a specific
artist or from a certain category). All queries are formulated
in SPARQL, which makes the query interface more flexible,
efficient, and also much more consistent.

The GAMA portal makes use of this in various ways, e.g.
through text fields for keyword search, filtering of result lists
by certain criteria such as genre or artist, or by application
of QbE aligned with other criteria. Browsing interfaces within
the GAMA portal allow for browsing media content based on
available metadata, e.g. in list interfaces.

The following sections describe the approach for result list
generation in the video and image QbE subsystems.

A. Video QbE subsystem

The approach applied for video QbE within the GAMA
system is shot-based. Features (or descriptors in the sense of
MPEG-7) are extracted and matched on the basis of video
shots (see Section V-A). A shot-based approach is chosen,
because, while for a complete video audiovisual characteristics
might severely change over time, within a single shot these
characteristics are much better defined in general.

Subsequently, also queries within the video QbE subsystem
are single shots. The result of a query is a list of videos
containing similar shots. As a distance measure for videos
the distance of the best matching shot can be assumed. As an
alternative, a voting approach based on all matching shots per
video in the best \( N \) matches is currently evaluated. Result
lists displayed to the user within the GAMA system are
video-based, so this approach is consistent with the common
approach for result presentation in the GAMA portal and video
QbE can be combined with other search criteria, such as genre
or artist.

The core functionality of the video QbE subsystem of the
content-based indexing service is the pre-generation of result
lists for every potential query, which is every indexed shot in
the sense of the GAMA system. Result lists are computed
on the basis of all features (or descriptors) and distance
measures (see Section V). For a description of the approach
for combination of distance measures refer to Section VI-C.
These lists are exported to the central RDF repository for each
shot.

B. Image QbE subsystem

The core functionality of the image QbE subsystem of
the content-based indexing service, similar to the video QbE
subsystem, is the pre-generation of result lists per image in the
database. Result lists are computed on the basis of all features
(or descriptors) introduced in Section V applicable for still
images. For a description of the approach for combination
of distance measures refer to Section VI-C. These lists are
exported to the central RDF repository for each image.

C. Pre-filtering and combination of feature distances

Both QbE subsystems for image and video matching
are based on multiple features (descriptors in the sense of
MPEG-7) and feature distances. Note that especially for
MPEG-7 descriptors a search according to a single descrip-
tor typically requires a one-to-one comparison of the query
descriptor with all database descriptors.

The worst case estimation for the database size in GAMA
is one million shots for the video QbE subsystem and one
million images for the image QbE subsystem. That is, an
exhaustive search for all descriptors over the complete dataset
would not be feasible. To overcome this, TZI PictureFinder
[5] (see Section V-C) is applied as a fast pre-filtering solution
within both image and video QbE.

A large ordered subset of the \( N_1 \) \((N_1 = 5000)\) best matching images or shots is selected through
a query to the PictureFinder system; here, best matches refer
to the visual similarity of key frames extracted by the shot
boundary detection and key frame extraction module (see
Section V-A). Using PictureFinder this ordered subset can be
computed in approximately 250ms on standard hardware for a
database of one million images/shots. For all images/shots in
this set a weighted combination of normalised distances for all
features is then computed and the set is reordered according
to this distance. The best matching \( N_2 \) \((e.g. N_2 = 100)\)
images/shots form the ordered set of results for a query.

VII. PRACTICAL EXPERIENCES

As the GAMA system is still in the implemetation phase, we
cannot report on experiences with the complete system online
at this point. Currently most of the underlying infrastructure
is already integrated and we are working on the design
and implementation of the GAMA portal that makes use of
the advanced search infrastucture for metadata described in
this document. The following sections list experiences and
problems encountered so far.

A. Heterogenity of data from the archives and harmonisation
issues

While integrating metadata coming from various databases,
we found out that the metadata are very heterogeneous in
terms of encoding systems, abbreviations, transliteration on
names, different spellings, same name variations, etc. This
usually is due to language differences and typing mistakes.
The expected results for a user query should be all the ones
fitting indeed the query, no matter how it is stored in the

The expected results for a user query should be all the ones
fitting indeed the query, no matter how it is stored in the

The expected results for a user query should be all the ones
fitting indeed the query, no matter how it is stored in the
different databases. If a user makes a query for the name of a person, for example, the expected results should be those which concern this person, no matter if the name is stored in the databases with misspellings, in different languages, with abbreviations, or otherwise. In order to solve this problem, a system that allows to “harmonize” the data stored in the databases is currently under development. The system will be based on approximate string matching algorithms and its aim would be to create a relation between names of persons and their spelling variations (considering possible spelling errors), but also of names of persons and art groups. This process is referred to as “harmonisation”.

B. Technical issues

Most content providers are not able or willing to host and maintain a complicated technical infrastructure (also see Section III). Therefore the GAMA system is designed in a quite centralised way, so that no software components need to be hosted on the side of the content providers. While such a system can be better maintained, there are also drawbacks. As content-based feature extraction, transcoding for streaming, and also the generation of media excerpts, such as thumbnails, is implemented in a central service, all media from the archives need to be transferred to the central servers in order to be processed. One the other hand some of the content providers have a quite limited bandwidth.

Another technical issue is the reuse of existing software components. GAMA is mainly an integration project. That is, most of the software components, like the feature extraction modules, are part of the content-based indexing service, and have been developed in completely different contexts and do not seamlessly fit each other.

C. Legal issues

As there are quite different types of content providers in the GAMA consortium ranging from festivals and small archives to even commercial distributors, we have to deal with varying legal conditions regarding access to the media art content. While some providers grant full access to media content without any restrictions others do not even allow access to media files at all. Within the GAMA system access to media files is needed for content-based analysis and to generate excerpts of media content (e.g. thumbnails) to enhance the portal. For content providers not allowing access to media files, obviously no content-based search functionality can be provided. Furthermore, within content-based analysis we extract representative frames for display in the portal and also generate a 30 seconds preview and a full flash-based streaming version of video artwork. Some content providers have no restrictions while others allow only a limited number of thumbnails to be displayed, and yet others even allow nothing to be shown. For streaming videos and previews the situation is similar.

Another legal issue is direct access to media and database servers on the side of the content providers. Again, some content providers allow this while others do not, mostly for legal reasons. Ideally the content-based indexing service would automatically download all media, execute content-based analysis of media, and then delete the media files from the central server. Similarly the database adaptors would directly connect to the archives of the databases, and collect and map relevant metadata. In order to deal with the restrictions above we had to offer other solutions, allowing content partner to upload (partial) database dumps and media files instead (also see Section III).

VIII. Conclusion

This paper described the approach for the enrichment of media content with metadata applied in the context of the GAMA project. The focus of this paper is to describe all components dealing with metadata in the GAMA architecture, which can be regarded as the “GAMA metadata engine”. This covers the centralised accumulation of metadata through database adaptors as well as the content-based analysis of media content by a central service. As has been shown, the centralised accumulation of available metadata from the archives and further enrichment of media content with audiovisual descriptions from content-based indexing will significantly improve the searchability of the underlying content.

ACKNOWLEDGMENTS

The GAMA project is funded under the eContentplus programme of the European Commission. The authors would like to thank the whole GAMA consortium.

REFERENCES