Automating Annotation of Media with Linked Data Workflows

Automating Annotation of Media with Linked Data Workflows

Thomas Wilmering, György Fazekas, Simon Dixon
Centre for Digital Music
Queen Mary University of London
London, E1 4NS
{t.wilmering, g.fazekas, s.e.dixon}@qmul.ac.uk

Sean Bechhofer
School of Computer Science
University of Manchester
Manchester, M13 9PL
sean.bechhofer@manchester.ac.uk

Kevin Page
Oxford e-Research Centre
University of Oxford
Oxford, OX1 3QG
kevin.page@oerc.ox.ac.uk

ABSTRACT

Computational feature extraction provides one means of gathering structured analytic metadata for large media collections. We demonstrate a suite of tools we have developed that automate the process of feature extraction from audio in the Internet Archive. The system constructs an RDF description of the analysis workflow and results which is then reconciled and combined with Linked Data about the recorded performance. This Linked Data and provenance information provides the bridging information necessary to employ analytic output in the generation of structured metadata for the underlying media files, with all data published within the same description framework.

Categories and Subject Descriptors

H.2 [**Database Management**]: Systems

General Terms

Multimedia databases

Keywords

audio metadata, feature extraction, linked data, semantic web

1. INTRODUCTION

Large online media libraries and the growing need for structured descriptions of these resources present several challenges in their design and implementation, including that the manual annotation of media resources, creation of linked data, and its reconciliation with other metadata sources can be extremely time consuming. There are no generic solutions for the automation of such analyses and for linking content derived metadata with editorial metadata; conversely ad-hoc solutions do not typically capture

the provenance of the computational workflow. This makes it difficult to subsequently identify errors in the procedure and to keep track of changes in the analysis data, e.g. due to changes of feature extraction algorithms. Furthermore, it complicates the reproduction of the data for research purposes or validation. These factors highlight the need for development of tools and approaches that can lower the cost of creating structured descriptions of large media resources, including automated generation of linked data and the means for using media feature extraction algorithms to support of its creation.

The Live Music Archive (LMA)[1] is a growing collection of over 100,000 live recordings of concerts, mainly in rock genres, made openly available with the permission of the artists concerned. Audio files are available in a variety of formats and each recording is accompanied by basic unstructured metadata describing information including dates, venues, set lists and the source of the audio files. In earlier work we introduced a linked data service that publishes metadata from the LMA, translated and restructured in an RDF form [1], and adding information asserting mappings to other collections such as MusicBrainz, GeoNames and last.fm. We have developed tools, described in this paper, to supplement the performance metadata with a computational analysis of the corresponding LMA audio data within the CALMA (Computational Analysis of the Live Music Archive) project [2]. Applications of our framework include feature derived structured annotation of the media files, enhanced search facilities for the LMA, and hybrid recommendation systems for professional users like radio producers, where we found that linking content and context is crucial to achieve user satisfaction [3].

2. COMPUTATIONAL ANALYSIS

We perform the computational analysis of the LMA audio data using Sonic Annotator[2], deploying a suite of audio feature extraction algorithms on multi-core compute servers. Sonic Annotator is able to output results in RDF, using concepts from the Music Ontology framework[3] and identi-

---

[1] https://archive.org/details/etree/
[2] http://www.vamp-plugins.org/sonic-annotator/
[3] http://musicontology.com/

fying audio files and associated timelines and events using document-local URIs. The set of features extracted from the audio material comprises of 41 features, including estimated boundaries of structurally consistent segments, chord times and labels, as well as lower level features, e.g. standard deviation, spectral inharmonicity and spectral centroid. The complete list of algorithms used for creating the metadata content can be found in [2].

The analytic and metadata generation workflow – from the retrieval of the audio material to the generation of linked data describing the audio features and audio features extraction process – is illustrated in Figure 1, and implemented using Python scripts. Losslessly compressed audio files retrieved from the LMA are converted to standard WAV PCM before we invoke the audio feature extraction. We store the results unchanged in individual *blobs*, i.e. compressed files, in order to minimise space requirements and preserve the integrity of the computational output. In addition, we generate RDF data from the computational output that allows provenance querying of the analysis processes and retrieval of analytic results.
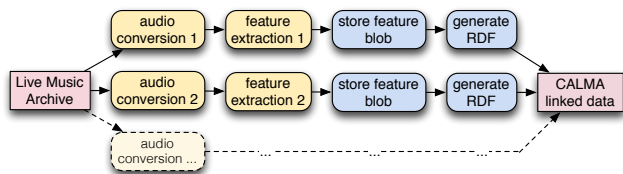

Figure 1: Processing pipeline for parallel audio feature extraction and linked data generation.

3. LINKED DATA GENERATION

For each audio track we create several resources to facilitate publishing the results of computational analyses, link them to editorial metadata resources (see Figure 2), provide provenance of computation and enable resource discovery:

analyses – overview of available features for a given track.
analysis_<hash> – provenance for an individual analysis.
analysis_blob_<hash>.tar.bz2 – compressed Sonic Annotator feature extraction output (RDF).
analysis_blob_side_<hash> – relates global URIs to local URIs within the blob (e.g. signals, timelines), which allows for querying uncompressed processing results.
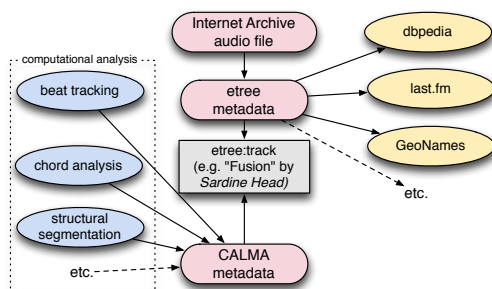

Figure 2: CALMA linked data generation and its connection to etree and the Internet Archive.

4. PROVENANCE

In order to fully describe the analysis performed – for self-description, reuse, reconciliation and repeatability – we provide provenance data about the media analysis algorithms applied in the metadata creation process, including software versions and parameters of the feature extraction processes. We use the PROV ontology[4] to describe software agents and activities, including details about the usage of these agents, such as commands used for their control. Figure 3 shows some of the provenance information for an individual audio feature resource.
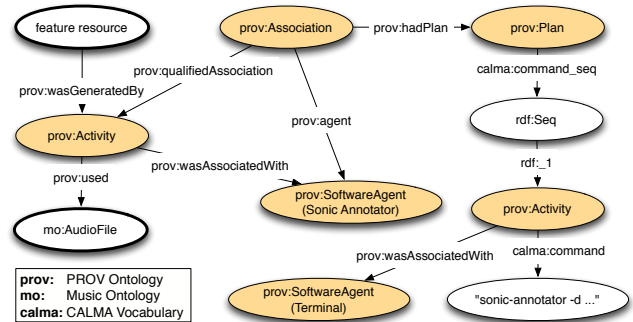

Figure 3: Provenance information of an audio features resource.

5. FUTURE WORK

Future work includes mapping feature semantics to higher level in-content concepts more directly comprehensible by end users, generalising the toolset, and creating plugins for the RDF generator to allow other computational analysis environments and media types to be supported.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] S. Bechhofer, K. Page, and D. De Roure. Hello Cleveland! Linked Data Publication of Live Music Archives. In *Proceedings of WIAMIS, 14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services*, 2013.

[2] S. Bechhofer, S.Dixon, G. Fazekas, T. Wilmering, and K. Page. Computational analysis of the live music archive. *Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR 2014)*, 2014.

[3] G. Fazekas, M. Barthet, and M. Sandler. The BBC desktop jukebox music recommendation system: a large-scale trial with professional users. *in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 15–19, July, San Jose, CA, USA*, 2013.

---

[4]http://www.w3.org/TR/prov-o/