The Networked Environment for Music Analysis (NEMA)

Kris West*, Amit Kumar†, Andrew Shirk†, J. Stephen Downie†, Andreas Ehmann† and Mert Bay†
*Widget Works Ltd
Norfolk, UK
Email: kris.west@gmail.com
†IMIRSEL, The Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
Champaign, IL, USA
Email: mirproject@lists.lis.illinois.edu

***Abstract*—Conducting valid comparative evaluations of techniques in the field of Music Information Retrieval (MIR) presents particular challenges to MIR researchers due to issues of copyright and data sharing. Further, the interdisciplinary nature of MIR research and multi-faceted nature of human music perception make the sharing and reuse of techniques and implementations for particular facets of music perception and music information retrieval tasks highly desirable. In addition the field makes use of a diverse range of file formats, software environments and toolkits for extracting, encoding and accessing MIR data and services, making reuse extremely challenging.**

The NEMA project aims to provide the MIR field with a high-quality, secure and extensible workflow environment to facilitate: computation over remote audio and resource collections; optimal code reuse, interoperability between data formats and types, sharing and dissemination; standardised, high-quality evaluation procedures; and the encoding of metadata, data and results in a format suitable for distributed systems.

I. Introduction

Unlike many other fields of multimedia research, Music Information Retrieval (MIR) suffers from extreme issues of data sharing and comparative evaluation. Owing to the current climate in the music industry, music collections cannot be freely shared between researchers (for example, those drawn from the set of western commercial music). Further, the purchase of licenses to the constituent tracks and reconstruction of a collection is often prohibitively expensive.

One solution to this problem is the use of creative commons collections, particularly those with research friendly licensing, such as the Magnatune collection [1]. However, such collections are often relatively small scale and they rapidly become well known to the research community (risking the over-fitting of multiple techniques and models to those collections). Further, multiple authors [2], [3] have reported significant (and common within the field) experimental errors introduced by poorly chosen experiment criteria. Hence, experiments must be both carefully constructed and exactly repeatable to facilitate valid statistical comparisons between techniques.

A. The Music Information Retrieval Evaluation eXchange

These challenges led to establishment of an annual 'Music Information Retrieval Evaluation eXchange' or MIREX [4]. MIREX is based on the TREC approach [5], where algorithms and applications are submitted by the MIR community to up to 26 separate, community-defined tasks and evaluated using standardised queries, collections and evaluation measures. The large number of tasks and datasets involved necessitates the sending of the code for each submission to the data, rather than the reverse (as is the case in most other fields).

Since the establishment of MIREX in 2005 (and its predecessor, the 'Audio Description Contest'), the annual evaluation campaign has been a strong driver of research in MIR by encouraging the community to clearly define tasks, datasets, research goals and standardised evaluation procedure that provide detailed and statistically valid comparisons of techniques.

Given that MIREX submissions must be delivered to the MIREX team to execute, the running of a MIREX evaluation is no small job. In 2009 alone, IMIRSEL received submissions from 138 individuals, in 26 tasks, generating a total of 289 evaluation runs. Since 2005, nearly 800 formal evaluation runs have been conducted and over 4Tb of audio files maintained for use in the various evaluation tasks.

B. Interoperability and Sharing

MIREX has so far failed to facilitate easy cross-pollination of research systems and code. Owing to the relative wealth of programming languages, software frameworks and toolkits used or developed for MIR research, such collaborations are often challenging as there exist few standards for encoding the wide range of music informatics data. Where standards do exist (such as the Music, Audio Feature and other ontologies developed by the OMRAS2 project for encoding MIR data in RDF [6]) uptake is as yet relatively low. Further, although MIREX provides a snapshot of the state-of-the-art in each MIREX task, it does not facilitate access to a wide range of published MIR techniques for less technical members of the MIR community such as Musicologists and Library Information Science researchers (the natural consumers of many MIR techniques). Hence, the NEMA project was established to address these issues.

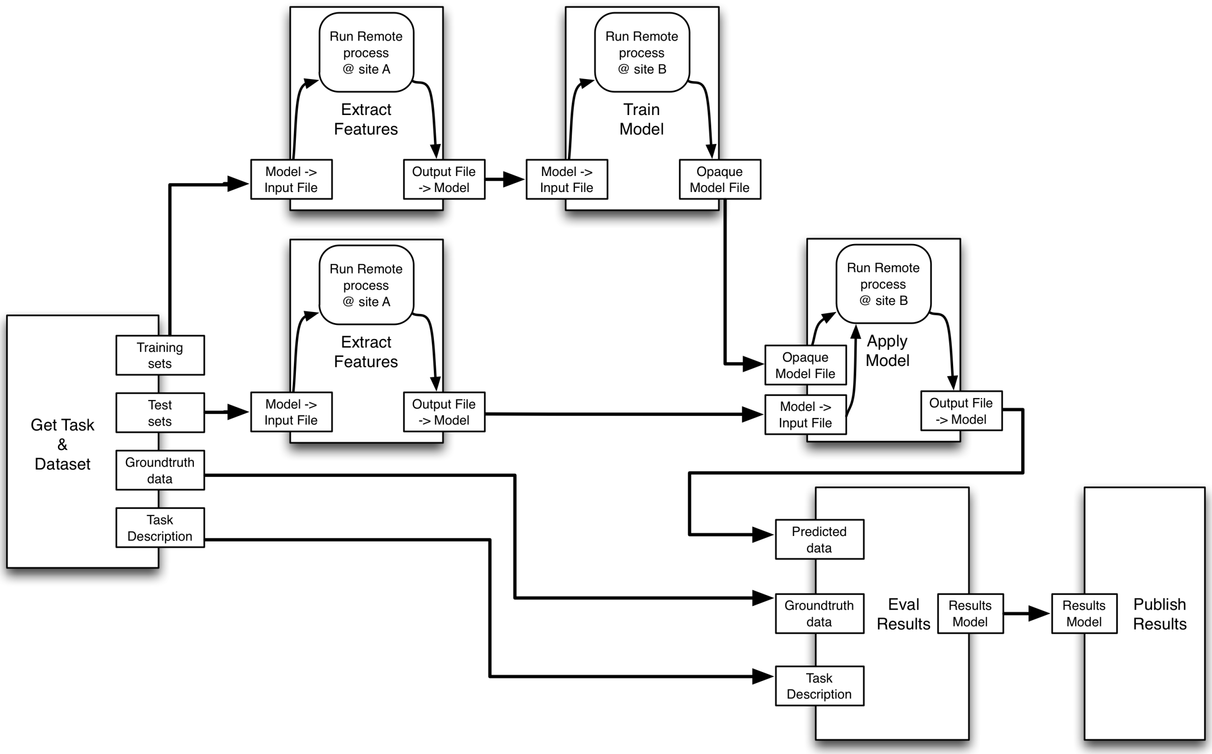

Fig. 1. An example of a NEMA implementation of an experiment workflow

C. Goals

The stated goals of the NEMA project include the establishment of a computational infrastructure to facilitate submission of MIR compute tasks against music collections and their metadata, interoperability between arbitrary systems for well-defined MIR tasks, standardised and automated submission/evaluation and the exposure of metadata and results in a standardised form suitable for integration with a full linked data architecture for MIR. Further, NEMA needs to be able to support the distributed execution of tasks against meta-collections of tracks drawn from multiple sites, without moving the content from those sites.

Though the NEMA project is not intended to replace MIREX, implementing support for the hosting of MIREX offers an excellent starting point for a field-wide computational infrastructure and interoperability as MIREX defines standardised tasks, collections and low-level file formats that are already supported by many softwares from the research community.

II. Implementation

By using a workflow approach to the definition of tasks that can be performed on the NEMA service, those tasks are broken down into discrete units or components which may be independently scheduled and distributed to satisfy collection access restrictions or software/architecture dependencies. For example in the current implementation, a user of the NEMA service would be able to conduct an experiment in which:

- a task and dataset is selected, drawn from of a collection of tracks held at a remote site,
- the dataset is analysed by Matlab code sent to the remote site and the resulting descriptors are imported into a NEMA data model (see II A) and transmitted to the home site,
- one or more machine learning frameworks are launched on a specified Virtual Machine infrastructure to train a model,
- the model is used to predict a metadata field or transcription for each track and the predictions are also imported into the NEMA data model,
- the results of the experiment are evaluated by MIR evaluation components provided by the NEMA project,
- finally the predictions and evaluation data are exposed by the NEMA service.

A diagrammatic overview of the workflow for this experiment is given in figure 1.

A. The NEMA Data Model

The NEMA infrastructure includes the definition of an extensible data model and data structures that encode data relating to all aspects of music, music informatics and MIR evaluation that have been encountered at MIREX, in the jMIR framework [8] or in the MIR related ontologies developed by the OMRAS2 project [6]. This data model and associated infrastructure facilitate interoperability between diverse components used on the NEMA service by users. User submitted code run by the NEMA service is wrapped in a component harness that uses contextual information about both the experiment and component code to move data between the NEMA data model and suitable file formats.

B. Infrastructure and Workflow Implementation

NEMA uses Meandre[1] to implement its worklow processes. Meandre is a flow execution environment that provides basic infrastructure for data-intensive computation. It provides a high-level language to describe flows and both a multicore and distributed execution environment based on a service-oriented paradigm. The Meandre Infrastructure provides a plugin model similar to the servlets in a Java Enterprise Enviornment where custom functionality is implemented.

The NEMA service leverages a Virtual Machine Infrastructure for the distributed execution of component code, triggered from within a Meandre component. Hence, workflows may be constructed that execute parts of a particular task at different sites and under differing architectures.

Throughout the hosting of five years of MIREX evaluations, a wide variety of dependency requirements have been encountered. Many of these dependencies are satisfied in the by our standard Virtual Machine Image. However, there are always special cases, including platform dependencies on Windows operating systems and non-standard libraries. Hence, a range of VM images are maintained to run user submitted code, wrapped as Meandre Components. In the future NEMA may provide a user customisable VM container system that will be exposed to the end user allowing them to login to an existing image, modify it for their use and save it for later reuse. Obviously, such modified VMs running potentially unknown code must be suitably sand-boxed.

The NEMA infrastructure currently supports user submitted and third party application components that execute Java, Matlab and arbitrary binary executables and will support VAMP API plugins [9] for music analysis in the future. This will provide an immediate path to the integration of many MIR softwares and provides a model for building hosts for further plugin environments. A diagrammatic overview of the NEMA service layers is given in figure 2.

1) Exposing Metadata, Predictions, Results and Provenance: The NEMA service infrastructure provides each workflow execution persistent storage of logs, execution information / provenance, predictions and evaluation results. This data may be published by a user within NEMA so that it maybe accessed by the community.

The design of the linked data system or *repository* for the NEMA project is beyond the scope of this paper. However, we store provenance and the results of a task execution to a Java Content Repository implementation [2]. JCR allows us to store arbitrary binary data along with the metadata and exposes multiple access API's including a WebDAV layer which will be used to provide on-demand processing required

[1]http://seasr.org/meandre/
[2]http://jackrabbit.apache.org/

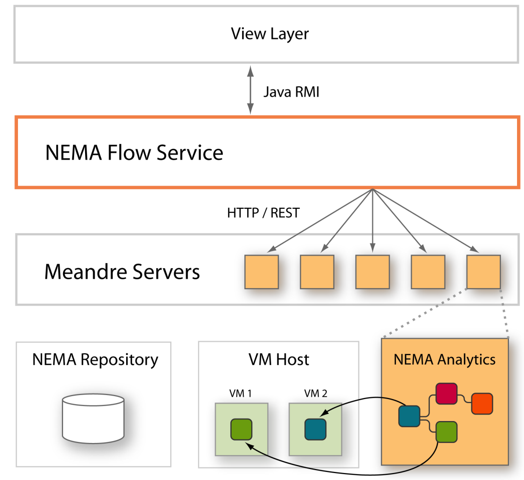

Fig. 2. An overview of the NEMA service infrastructure

to satisfy queries submitted via the repository endpoint (i.e. extracting and/or predicting metadata for content that is not already known).

C. Resource Resolution, Security and Attribution

Given that security is a prime concern for any NEMA-hosted collections, it is vital that the analysis code is sent to the data and that result data is only returned in a controlled fashion in which the transmission of the original audio content (or a reconstructable form thereof) is prevented. The model for the extraction of MIR information on remote collections used in NEMA, which protects the NEMA-hosted music collections, is inspired by the On-demand Metadata Extraction Network (OMEN) [11] developed at the Schulich School of Music, McGill University.

Format dependencies means that multiple different encodings of the content are required to be maintained or conversion facilities provided. The ability to represent multiple sources for the content facilitates better access to the resource for example, NEMA might not be able to provide streaming access to the controlled content for auditioning but could refer to a publicly or commercially accessible version of the content.

The security of code submitted to the NEMA service is a prime concern, as these components may or may not be under IP restriction or may not be ready for wide-spread distribution. Hence, the NEMA service allows users contributing component code to control the level of access of other users to their components (implementing a 'some rights reserved' sharing model that may help promote academic work and experimentation that would otherwise be difficult or impossible). Additionally, full author attribution information and citation information is captured with each component profile in order to facilitate correct attribution in cases of reuse.

III. The NEMA Service and Workflow Sharing

The existing approach to the running of MIREX evaluations involves the manual execution of diverse pieces of code by a half dozen research students over a two month period and is therefore considered expensive in terms of human resources. By hosting and exposing the MIREX evaluations as workflow templates through NEMA's service infrastructure, the configuration, execution and logging of the task workflow is exposed to the user. This enables the user to solve any issues they encounter in order to produce a successful run, including corrections/alterations to their uploaded codebase or configuration. Hence, this may be considered a significant advance on the current manual process. Further, experimentation need not be confined to an annual evaluation and will eventually be accessible year-round to the MIR research community.

The workflow sharing approach designed for NEMA has been modelled after the myExperiment Virtual Research Environment [12] and directly addresses many of the same issues. It is expected that the NEMA service will integrate with myExperiment at some stage in the future. This approach facilitates the sharing and reuse of experiment setups and partially or fully-specified implementations of those experiments. For example, many MIR tasks may be defined in terms of a workflow template with a component that provides an input dataset and task description (e.g. classify music by genre), a number of required but unspecified process components (e.g. perform feature extraction, train classifier, apply classifier to test dataset) and an evaluation component. A user can configure components to perform each of the processes specified in the workflow template and execute the configured workflow to perform the task and receive evaluation results. Hence, a researcher will be able to publish an experiment definition (a workflow template) and either fully configured workflows or simply computed results representing experiments in published papers or theses, allowing other researchers to accurately compare their work to or build upon the results.

Further, if users choose to publish their components, NEMA can drive the cross-pollination of research mitigating the need to re-implement techniques or learn additional languages/frameworks/applications. This lowers the bar of entry into field, allowing researchers to focus on their particular area of research, rather than supporting technologies. It is hoped that NEMA will also facilitate both Musicological research in lower time-scales, skill-sets and levels of effort than was previously possible and form a ideal teaching tool for use in many fields including Musicology, Machine-learning, Library Information Science and MIR.

Finally, the provision of automated evaluation tools and access to large-scale analysis collections should facilitate (as MIREX has done in the past) better analysis, reporting and dissemination of results in the field of MIR than would otherwise be possible.

The evaluation tools and IO/Interoperability framework developed to implement the NEMA service is available as an open-source toolkit for MIR experimentation and software building. It is hoped that this will be a significant resource to the community in and of itself[3].

[3]source code and documentation for the NEMA project infrastructure is available from google code: http://code.google.com/p/nemadiy

IV. CONCLUSION

The NEMA project attempts to provide a workflow architecture for the MIR domain that uses disparate systems, frameworks and data models. The self-service approach lowers the bar for the participants of MIREX, and will allow them to use several standard MIR libraries such as jMIR [8], Sonic Annotator and the VAMP API [9] in conjunction with their own research.

ACKNOWLEDGMENT

The NEMA project is funded by the Andrew W. Mellon foundation. The authors would also like to acknowledge the contributions of the NEMA project team at the IMIRSEL laboratory at UIUC and the project partners from the Universities of McGill, Southampton, Waikato, Goldsmiths University of London and Queen Mary's University of London.

REFERENCES

[1] J. Buckman, "Magnatune: Mp3 music and music licensing," http://magnatune.com, April 2006.

[2] Y. Kim, D. Williamson, and S. Pilli, "Towards quantifying the album effect in artist identification," in *Proceedings of ISMIR 2006 Seventh International Conference on Music Information Retrieval*, September 2006.

[3] E. Pampalk, "Computational models of music similarity and their application in music information retrieval," Ph.D. dissertation, Johannes Kepler University, Linz, March 2006.

[4] J. S. Downie, "The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research," *Acoustical Science and Technology*, vol. 29, no. 4, pp. 247–255, 2008.

[5] J. Tague-Sutcliffe and J. Blustein, "A Statistical Analysis of the TREC-3 Data," *Overview of the Third Text Retrieval Conference (Trec-3)*, 1995.

[6] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson, "The music ontology," in *Proceedings of the International Conference on Music Information Retrieval*, 2007, pp. 417–422.

[7] R. Lee and S. Seligman, *The JNDI API tutorial and reference: building directory-enabled Java applications*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2000.

[8] C. McKay and I. Fujinaga, "jMIR: Tools for automatic music classification," in *Proceedings of the International Computer Music Conference*, 2009.

[9] C. Cannam, C. Landone, M. Sandler, and J. Bello, "The sonic visualiser: A visualisation platform for semantic descriptors from musical signals," in *Proceedings of the International Conference on Music Information Retrieval*. Citeseer, 2006, pp. 324–7.

[10] I. Witten, S. Boddie, D. Bainbridge, and R. McNab, "Greenstone: a comprehensive open-source digital library software system," in *Proceedings of the fifth ACM conference on Digital libraries*. ACM New York, NY, USA, 2000, pp. 113–121.

[11] D. McEnnis, C. McKay, and I. Fujinaga, "Overview of OMEN," in *Proceedings of the International Conference on Music Information Retrieval*. Citeseer, 2006, pp. 7–12.

[12] D. De Roure, C. Goble, and R. Stevens, "Designing the myexperiment virtual research environment for the social sharing of workflows," in *Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*. IEEE Computer Society Washington, DC, USA, 2007, pp. 603–610.