A "Do-It-Yourself" Evaluation Service for Music Information Retrieval Systems

M. Cameron Jones Mert Bay J. Stephen Downie Andreas F. Ehmann
University of Illinois at Urbana-Champaign
501 E. Daniel St. Champaign, IL 61820 USA
{mjones2, mertbay, jdownie, aehmann}@uiuc.edu

Categories and Subject Descriptors
H.3.4 [**Performance Evaluation**]: Efficiency and Effectiveness

General Terms: Performance, Measurement, Experimentation, Standardization

Keywords Music Information Retrieval, Evaluation

1. EXTENDED ABSTRACT
This demonstration presents the Do-It-Yourself (DIY) web service of the Music Information Retrieval Evaluation eXchange (MIREX). As TREC does for text retrieval, MIREX provides standardized datasets and evaluation frameworks to evaluate Music Information Retrieval (MIR) systems and algorithms [1]. However, unlike TREC where participants are given the datasets and execute their code locally, MIREX data sets cannot be distributed due to copyright restrictions. In previous years, MIREX participants submitted systems to the International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL), where they were manually executed, and evaluated.

The MIREX DIY service (see a demo at http://music-ir.org/mirexdiy/) allows for the remote execution of black-box algorithms submitted by participants, and provides participants with real-time progress reports, debugging information, and evaluation results. The DIY service can be remotely controlled by the participants, allowing for continuous submission, evaluation, and improvement of algorithms, mitigating the intensive debugging, execution, and validation efforts previously required of IMIRSEL members [1].

2. SYSTEM OVERVIEW
The MIREX DIY service extends the Data-to-Knowledge Web Service (D2KWS) [3] and Music-to-Knowledge (M2K) libraries [2] to support the remote execution of participant-submitted MIR algorithms (Figure 1). Participants submit algorithms via a web interface to MIREX DIY. MIREX DIY can currently support compiled C/C++ or Java binaries, as well as Matlab, Perl, and Python scripts. Using either SOAP (Simple Object Access Protocol) or a web interface, participants then specify an itinerary in which to execute their algorithm. The itinerary contains information about the dataset(s) to be evaluated, the participant-specified algorithm, and the evaluation metric to be used. This information is passed to the D2KWS which queues the to-be-executed job. D2KWS distributes the queued jobs to a set of sandboxed and firewalled D2K servers. Participants can monitor a job's status and console output via SOAP or a web interface. Job results are stored in the D2KWS and are returned to the client. In the event of failure, exceptions and error messages are logged and returned to participants. Participants can revise algorithms and resubmit for future evaluation.

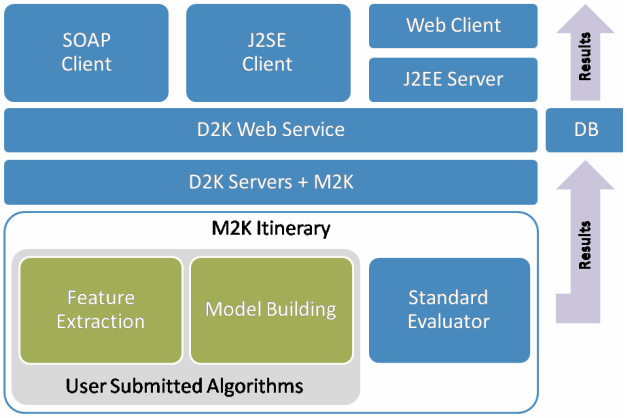

Figure 1. System Architecture.

3. CHALLENGES
Providing the MIREX DIY service to the MIR community raises several challenges. The current MIREX datasets comprise over 1TB of audio and symbolic music files. This raises challenges with respect to the computational complexity and storage requirements for executing submitted algorithms. For example, a single job can take days to execute and generate gigabytes of feature sets and models as output. Furthermore, as algorithms are executed unsupervised, MIREX DIY must be robust to malicious code, not only protected against attacks (both intentional and accidental), but also secured against theft of copyrighted content.

4. REFERENCES
[1] Downie, J. S. (2006). The Music Information Retrival Evaluation eXchange (MIREX), D-Lib Magazine, 12(12).

[2] Downie, J. S., Ehmann, A. F., Tcheng, D. K. (2005). Music-to-knowledge (M2K): a prototyping and evaluation environment for music information retrieval research. SIGIR 2005, 676.

[3] Shirk, A. (2004). D2K Web Service Design & Implementation, presented at NCSA CyberArchitecture Working Group, Available at http://algdocs.ncsa.uiuc.edu/PR-20040828-1.ppt

SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.
ACM 978-1-59593-597-7/07/0007.