

Evaluatron 6000: Collecting Music Relevance Judgments

Anatoliy A. Gruzd, J. Stephen Downie, M. Cameron Jones, Jin Ha Lee
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
+1 (217) 333-3280

{agruzd2, jdownie, mjones2, jinlee1}@uiuc.edu

Categories and Subject Descriptors

H.3.7 [Digital Libraries]: User issues

General Terms

Measurement, Performance, Human Factors.

Keywords

MIREX, Music Digital Libraries, Music Information Retrieval, Music Similarity.

1. INTRODUCTION

Human relevance judgments are critical in evaluating digital library collections and retrieval systems. Within the Music Digital Library (MDL) / Music Information Retrieval (MIR) community, there is a paucity of tools that allow researchers to systematically collect and analyze music similarity/relevance judgment data. To aid in the collection of ground truth similarity data, we developed a web-based system called the Evaluatron 6000 (E6K).

E6K was first deployed in support of the 2006 Music Information Retrieval Evaluation eXchange (MIREX) [1] to evaluate algorithms submitted to the “Audio Music Similarity and Retrieval” (AMS) and “Symbolic Melodic Similarity” (SMS) tasks [2]. In September 2006, the E6K was used to collect 7602 similarity judgments from 45 “graders”.

2. SYSTEM OVERVIEW

E6K collects relevance judgments for query-candidate pairs (QCP) generated by the submitted algorithms. These QCP serve as the primary unit of interaction in the system. Users can audition samples of queries and candidates for each QCP and are asked to input two similarity evaluations per QCP: 1) a BROAD category of similarity (i.e., Not Similar (NS), Somewhat Similar (SS), and Very Similar (VS)); and, 2) a FINE score between 0.0 (Least similar) and 10.0 (Most similar) (Figure 1).

2.1 Detailed System Architecture

E6K is built upon the “CMS Made Simple” open-source content management system [3] which both reduced the development time and simplified system management. As a web-based application, E6K adheres to a Client-Server model: the client consists of HTML, CSS and JavaScript; and, the server – PHP and MySQL. This web-based approach has the benefit of allowing graders to use the system from anywhere they have a browser and an Internet connection. E6K employs a popular web 2.0 programming technique referred to as AJAX (Asynchronous JavaScript and XML) [4] to save similarity/relevance judgments

and other interaction events in real time, allowing graders to leave the system and come back as they wish. As a side benefit, use of AJAX also prevents data loss in the event of unexpected service interruptions or system failures.

To ensure cross-browser/platform compatibility, E6K gives graders a choice of three audio players: Flash, Windows Media Player, and Quicktime. All players draw from a common set of QCP MP3 files. The E6K tracks and records all user-interactions with the system. For MIREX 2006, this consisted of 69,745 logged events.

This demonstration is intended to illustrate the major features of the E6K system, and to prompt discussions about its potential uses in other digital library evaluation contexts.

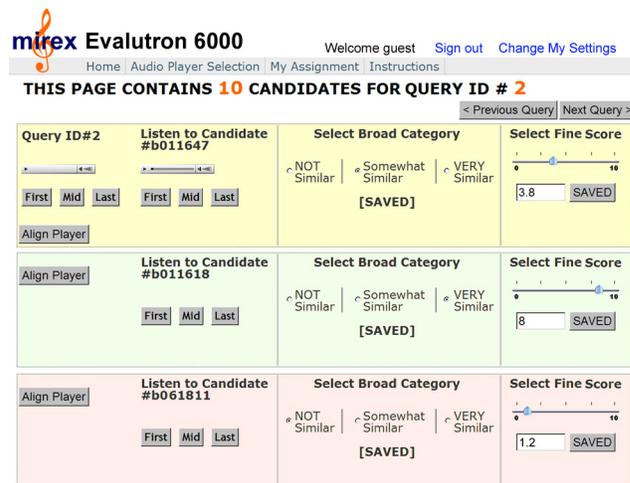


Figure 1. Screenshot of the Evaluatron 6000 interface.

3. ACKNOWLEDGMENTS

Special thanks to: The Andrew W. Mellon Foundation, the National Science Foundation (Grant No. NSF IIS-0327371), and the MIREX 2006 graders.

4. REFERENCES

- [1] Downie, J. S., West, K., Ehmann, A., and Vincent, E. The 2005 Music Information Retrieval Evaluation eXchange (MIREX 2005): Preliminary overview. In *Proceedings of the 6th Int'l Conference on Music Information Retrieval (ISMIR 2005)*, Queen Mary, UK, 2005, 320-323.
- [2] MIREX Wiki. Available at: <http://music-ir.org/mirexwiki/>.
- [3] CMS Made Simple <http://www.cmsmadesimple.org>.
- [4] Ajax [http://en.wikipedia.org/wiki/Ajax_\(programming\)](http://en.wikipedia.org/wiki/Ajax_(programming)).

Copyright is held by the author/owner(s).

JCDL '07, June 17–22, 2007, Vancouver, British Columbia, Canada.
ACM 978-1-59593-644-8/07/0006.