



BA:

Learning From User Data

This document describes the subject and the general time schedule of the bachelor thesis of Tobias Schlüter, beginning in the spring term 2012. Adaptations or changes can be agreed upon by the advisors.

Several factors, such as the growth of the Internet, peer-to-peer technologies, or the emergence of the compact media formats have changed the way people deal with music. Personal music collections have grown bigger, and, thanks to portable players and advances in storage technology, they can nowadays be accessed anywhere and anytime. The music collections accumulated by music lovers have reached sizes that make it hard to maintain an overview of the data by just browsing hierarchies of folders.

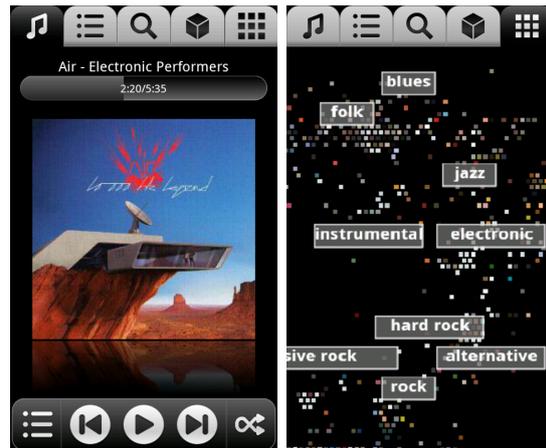
In our lab, we have developed jukefox (<http://www.jukefox.org>), a Music Player for the Android Mobile Platform that addresses these issues. An important ingredient of the application is a map of music that facilitates a multitude of similarity aware browsing and playlist creation interfaces.

Currently these smart features are based on user data from the online radio service *last.fm*. However, we have been collecting usage data from the jukefox users about their music libraries and their listening behaviors for several months. This led to a log of over 100 million entries of tens of thousands of users. The goal of this thesis is to use this data to replace the *last.fm*-based music similarity measure and to extract other recommendation-related information from it. To achieve this goal, Tobias will compare different embedding and recommendation methods and implement an algorithm that allows to place songs or artist in an Euclidean music similarity space and give personalized recommendations. He will then evaluate the quality of the similarity space and the recommendations in a user study or other ground truth data.

Requirements: Good programming skills (preferably in Java) and some creativity are advantageous. The student(s) should be able to work independently on the topic.

Contacts

- Samuel Welten: swelten@tik.ee.ethz.ch, ETZ G61.4



Detailed Project Outline

We denote the following primary tasks mandatory (on the right side you find a rough estimate for the time that we allocate to the respective task):

- Get used to the music similarity measure and jukefox (★)
- Compare different embedding and recommendation algorithms (★★)
- Implement an embedding that results in a Euclidian music similarity space (★★★)
- Implement a recommendation algorithm that is not computation intensive and works without having to connect to a central server for every recommendation (★★★)
- Implement an interface to query music similarity coordinates and recommendation information (★★)
- Evaluate the quality of the algorithms (★)
- Write a report documenting the development process and the final status of the application and discuss the findings. (★★)
- Prepare a presentation about the results of your work (★★)

The Students' Duties

- One meeting per week with the advisors to discuss current matters
- Regular check-ins into the provided *revision control system* (Subversion)
- A final presentation (15 min) of the work and results obtained in the thesis
- A final report (English or German), presenting work and results
- Independent working is expected
- A possibility to work in the ETZ is provided. It is also possible to work at home