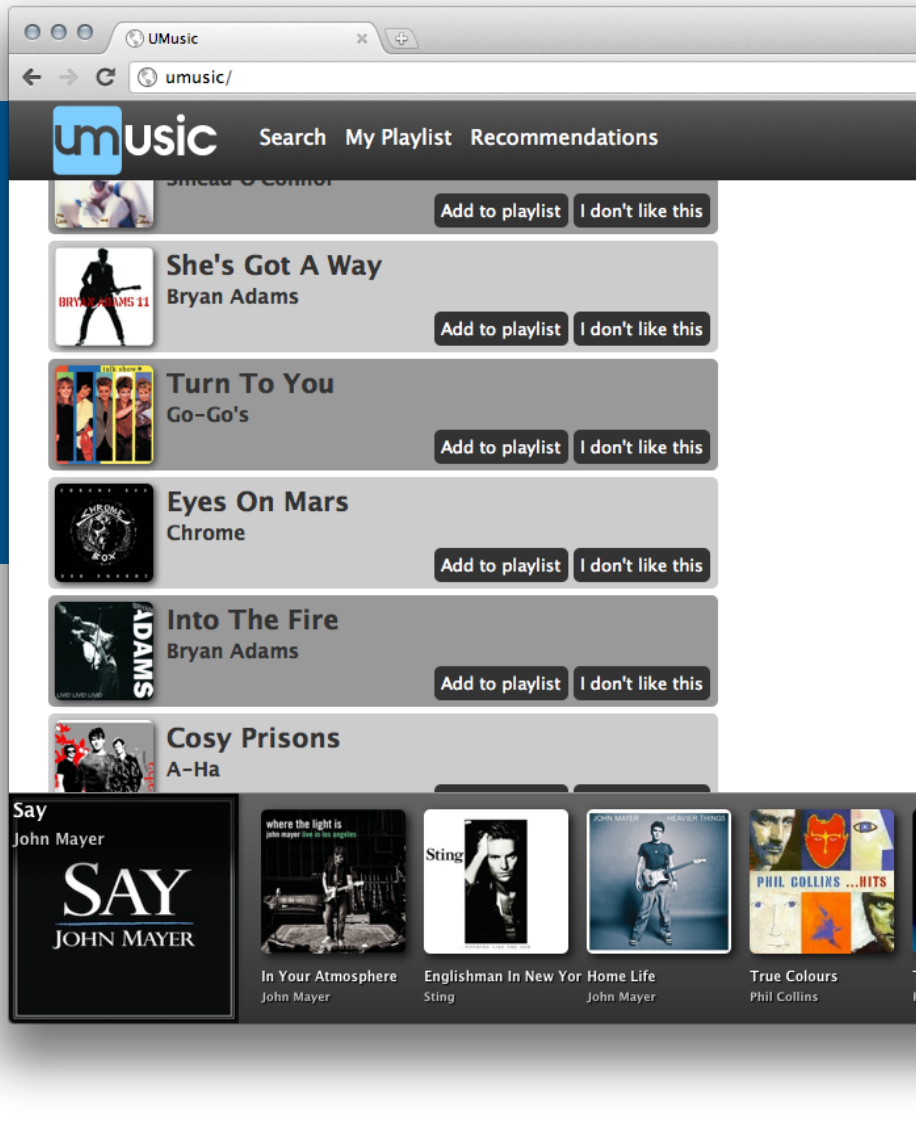
Project Information and Knowledge Engineering

Bojana Dumeljic
Berend Ottervanger
Hylke Visser

What is Umusic?

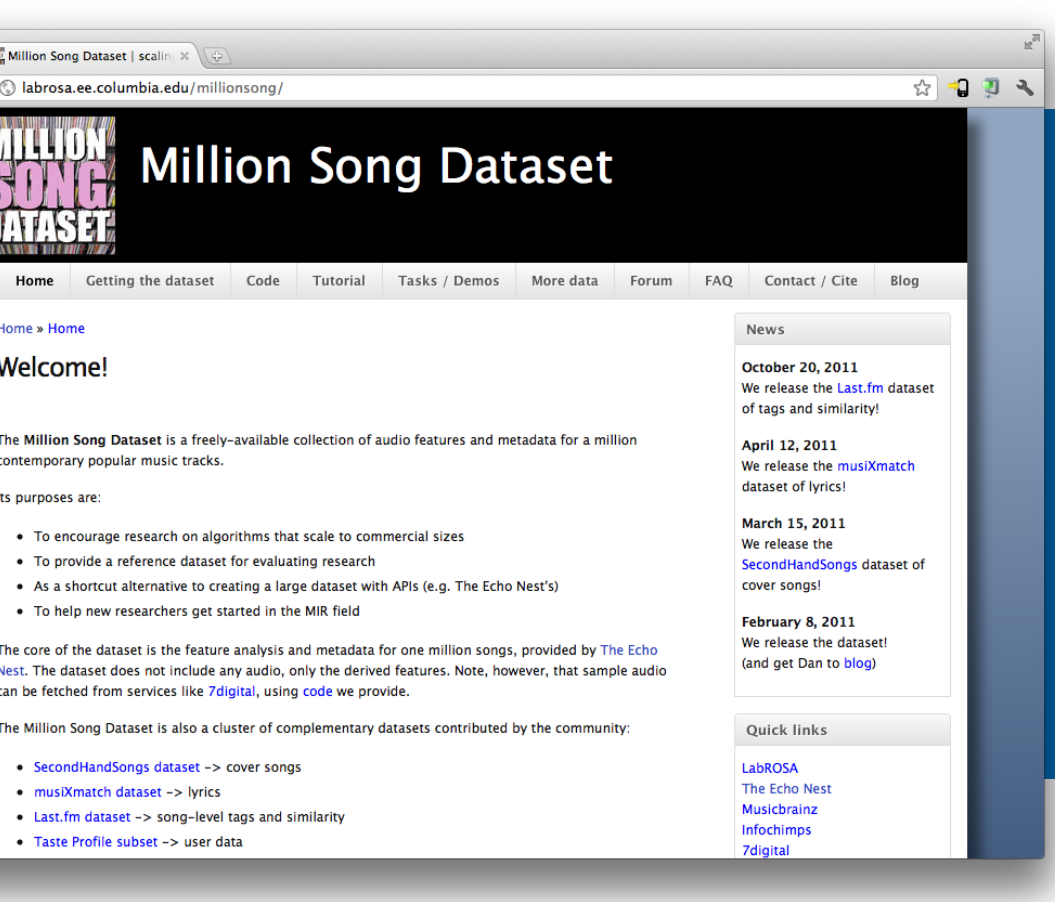
Umusic is a web based music player and recommender system. It lets users search for music and play it. The information the system gathers from the user actions is used to build a user profile and recommend other songs to the user, songs the user doesn't know yet.



Behind the scenes

The Umusic system is built with PHP. It uses Kohana, a PHP framework that provides a rich set of components for building web applications.

The client side uses jQuery to respond to user actions and Mustache to render the different parts of the application.



Datasets

The Umusic system uses two datasets. The largest dataset is the one containing the track metadata, Title, Artist name, Album and duration. This data is part of the Million Song Dataset, provided by the Columbia University.

The other dataset is a dataset of Last.fm tags associated to songs. This dataset contains the information we use to recommend songs to users.

The recommendation system

Every user action is saved in the database. These actions have a corresponding rating. The action “add to playlist” has a rating of +1, the action “remove from playlist” -1 etc. This information is periodically processed to update the user profiles.

We chose an item based recommendation system. Every user and evey track has a corresponding vector with “rating” values for a selection of tags. The majority of these tags represent music genres, but there are also tags describing a mood or artist or indicating whether a song is a live performance.

User ID	Track ID	Action
1	TRCCCYE12903CFF0E9	liked
1	TRCCCFH12903CEBC70	added
2	TRCCCYE12903CFF0E9	liked
2	TRCCCBJ128F4286E6F	liked
1	TRCCJTD12903CB62E6	liked
1	TRCCJKN128F426E04D	disliked
1	TRCCILA128F42BA535	removed
2	TRCCITF12903D009A3	disliked
2	TRCCHLU128F92FA064	added
2	TRCCHIV128F14A15C9	added

Processing the user actions

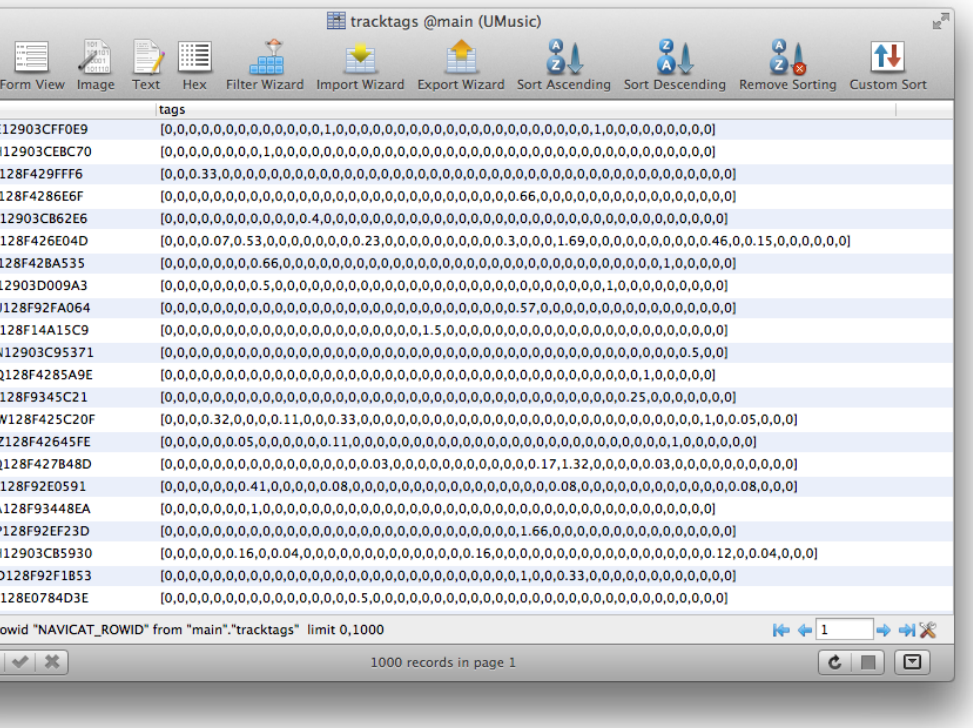
For each user the system fetches all performed actions from the database. Next, it creates a vector of the tags that are attatched to each track in this list, which is multiplied with the action rating. These vectors are then summed up and stored in the database.

User	Vector
1	[1,1,1,1,-1,-1,-1]
2	[1,1,1,1,0,0,0]

Track	Tag
TRCCCYE12903CFF0E9	rock
TRCCCYE12903CFF0E9	pop
TRCCCYE12903CFF0E9	singer-songwriter
TRCCCYE12903CFF0E9	90s
TRCCJKN128F426E04D	electronic
TRCCJKN128F426E04D	dance
TRCCJKN128F426E04D	dj

Track conversion

Recommending songs to users can only be done if users can be compared to tracks. Therefore we had to convert the information we have about each track to a vector similar to the user vector.
For each track we checked which of the tags that we selected are attached to the track. These results are stored in a vectorized format. The process of gathering and converting the track information is a very complicated and time consuming process. It took about six hours to calculate the vectors for all the songs in our database.



Comparison and recommendation

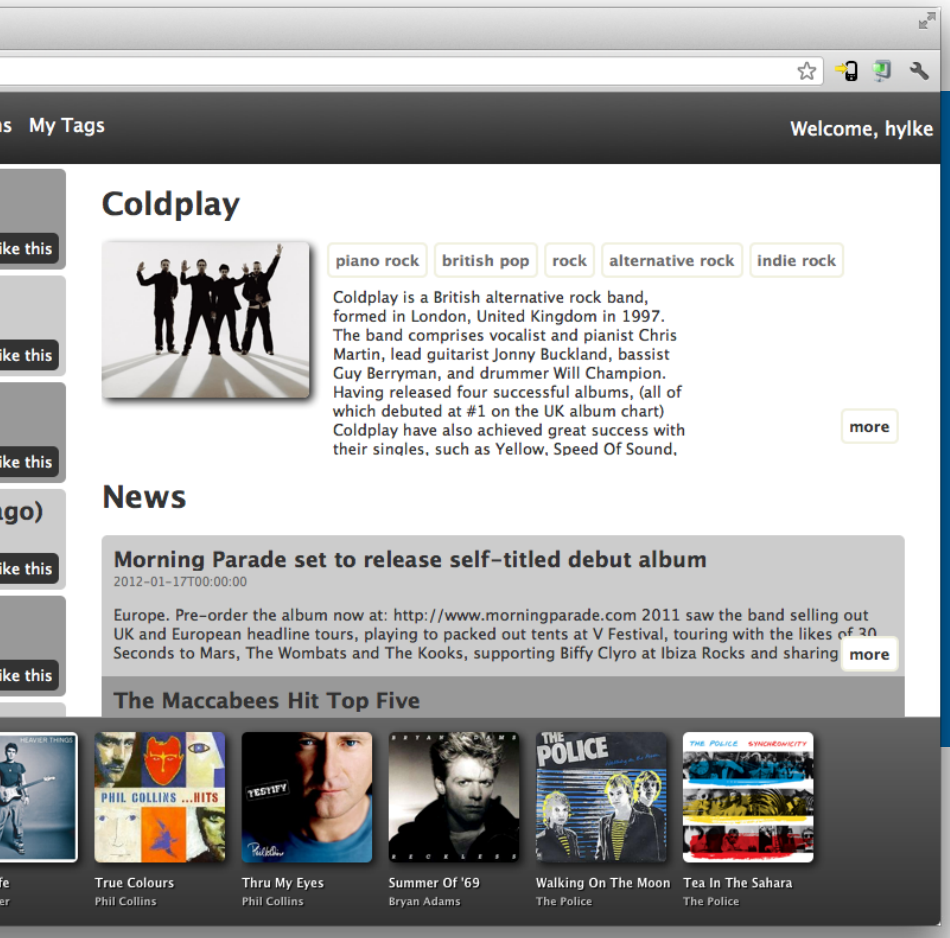
When all user and track vectors are calculated, we can begin with recommending music to users. Our first approach was to calculate the similarities for each track and return the 100 songs with the highest similarities. These similarities are calculated to compare the track vector to the user vector, using the cosine similarity function.
Unfortunately, going through the entire database and calculating similarities for each track had too much impact on the performance of our server. The second approach was to start calculating similarities and stop when we have found 100 tracks with a minimum similarity of 80%.

umusic

by Bojana Dumeljic, Berend Ottervanger and Hylke Visser

More than a recommender system

Umusic can do a lot more than just recommend music to users. We use the power of several publically available APIs to enrich the recommender system with extra features. By using the YouTube data and player APIs it is possible to stream the music in the user's playlist. The Last.fm API provides the location of album art for each song. The Umusic system downloads the images and stores them locally. This makes the system graphically appealing to the user. We use the Echo Nest API to get information, news and reviews about a certain artist.



Artist information

When a user clicks on a track in the playlist, recommendations list or search results, we display information, news articles and reviews about this artist.

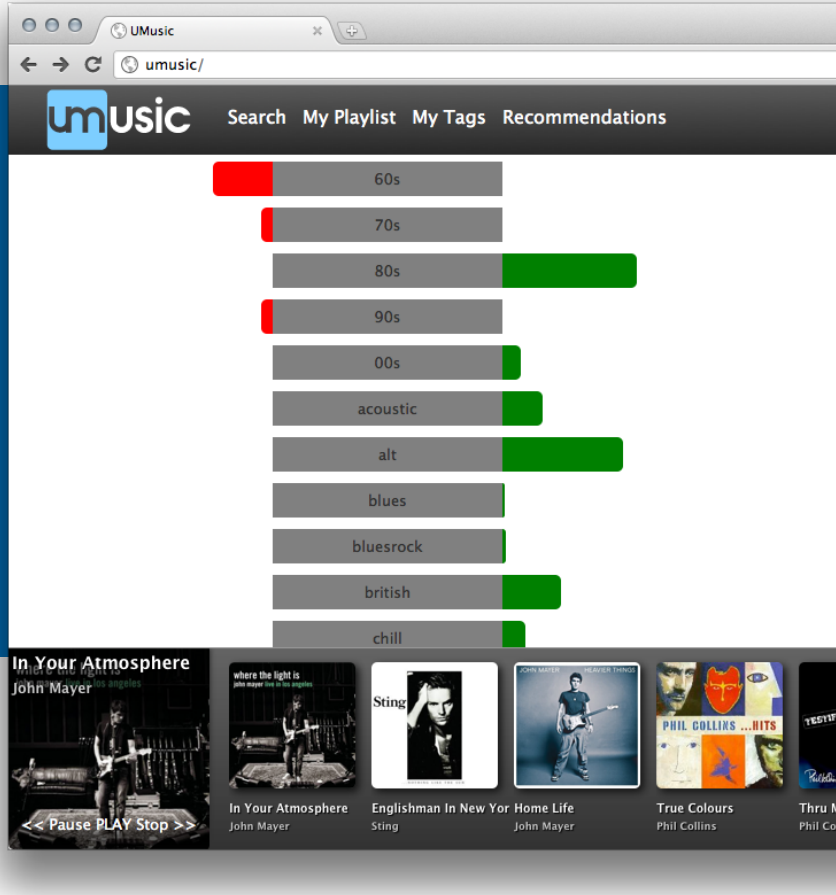
If a user wants to see more, he can click on the "more" link. This leads the user to the website where the article or review was originally posted.

YouTube integration

With the YouTube data and player APIs, we are able to stream music to the user. The data API provides the system with the ids of the songs in a user's playlist. Using the player API, we load these ids into the YouTube player and start playing them.

View your user profile

Users have the ability to see what their user profile looks like. The "My Tags" page gives an overview of the tags used in the Umusic system, along with a graphical representation of the user appreciation level for each tag.



umusic

by Bojana Dumeljic, Berend Ottervanger and Hylke Visser