DUNYA: A SYSTEM FOR BROWSING AUDIO MUSIC COLLECTIONS EXPLOITING CULTURAL CONTEXT

Alastair Porter, Mohamed Sordo, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{alastair.porter,mohamed.sordo,xavier.serra}@upf.edu

ABSTRACT

Music recommendation and discovery is an important MIR application with a strong impact in the music industry, but most music recommendation systems are still quite generic and without much musical knowledge. In this paper we present a web-based software application that lets users interact with an audio music collection through the use of musical concepts that are derived from a specific musical culture, in this case Carnatic music. The application includes a database containing information relevant to that music collection, such as audio recordings, editorial information, and metadata obtained from various sources. An analysis module extracts features from the audio recordings that are related to Carnatic music, which are then used to create musically meaningful relationships between all of the items in the database. The application displays the content of these items, allowing users to navigate through the collection by identifying and showing other information that is related to the currently viewed item, either by showing the relationships between them or by using culturally relevant similarity measures. The basic architecture and the design principles developed are reusable for other music collections with different characteristics.

1. INTRODUCTION

The research being performed within the CompMusic [14] project aims to provide new technologies, interfaces and navigation methods for browsing music collections from specific music cultures. In particular we are working with several non-western music traditions and are taking advantage of their specific characteristics to promote an approach in MIR that emphasises the use of domain knowledge in every step of the research project.

Our research goal is to extract musically meaningful descriptions from audio recordings by using as much contextual information as possible. From these descriptions we can develop similarity measures between all the entities that have been identified in a particular music collection.

Several music navigation and discovery systems have been proposed in the last few years [8, 9, 15]. Although differing in visualisation aspects, they all share the notion of computing audio similarity for music discovery. Other systems, like the *Sonic Visualiser* [2], are designed for visualising low-level and mid-level audio features. One of the main advantages of Sonic Visualiser is that it allows the integration of third party algorithms through the use of plug-ins. More recently, Goto et al. proposed *Songle* [3], a system that, among other things, allows users to actively listen to audio while visualising musical scene descriptors by engaging them in correcting estimation errors of the extracted audio features.

Some of these systems make use of related semantic information from the Internet and other sources, however none use cultural-specific information to build a comprehensive music listening experience.

In this paper we introduce Dunya, a system that allows users to explore a given music collection and discover musically relevant relationships between all the items that have a musical relevance. The name Dunya means world [1] in several languages of the cultures that we are studying. In Dunya we are emphasising the concepts of exploration and similarity based on cultural specificity and the design of the interface and of the system architecture emphasises this idea.

For this first version of Dunya we are focusing on Carnatic music, the classical music tradition of the south of India. In the next sections we will describe the architecture of the Dunya system, including the types and the sources of the information that we are using. We then describe the development of our interface for a music collection of the Carnatic music tradition.

2. ARCHITECTURE

This section presents the overall architecture of the Dunya system. Dunya consists of three main aspects (Figure 1). Firstly, raw data is gathered from many different sources. This data includes audio and other information about the music and musicians who performed it. The second aspect is the development of a data schema (a structured data representation) that specifically reflects the concepts that are found in a music tradition. The data schema allows us to store relationships and similarities between items in the database. The final part of Dunya is a graphical interface

[1] Originally coming from the Arabic language, Dunya means more precisely world in a metaphysical sense.

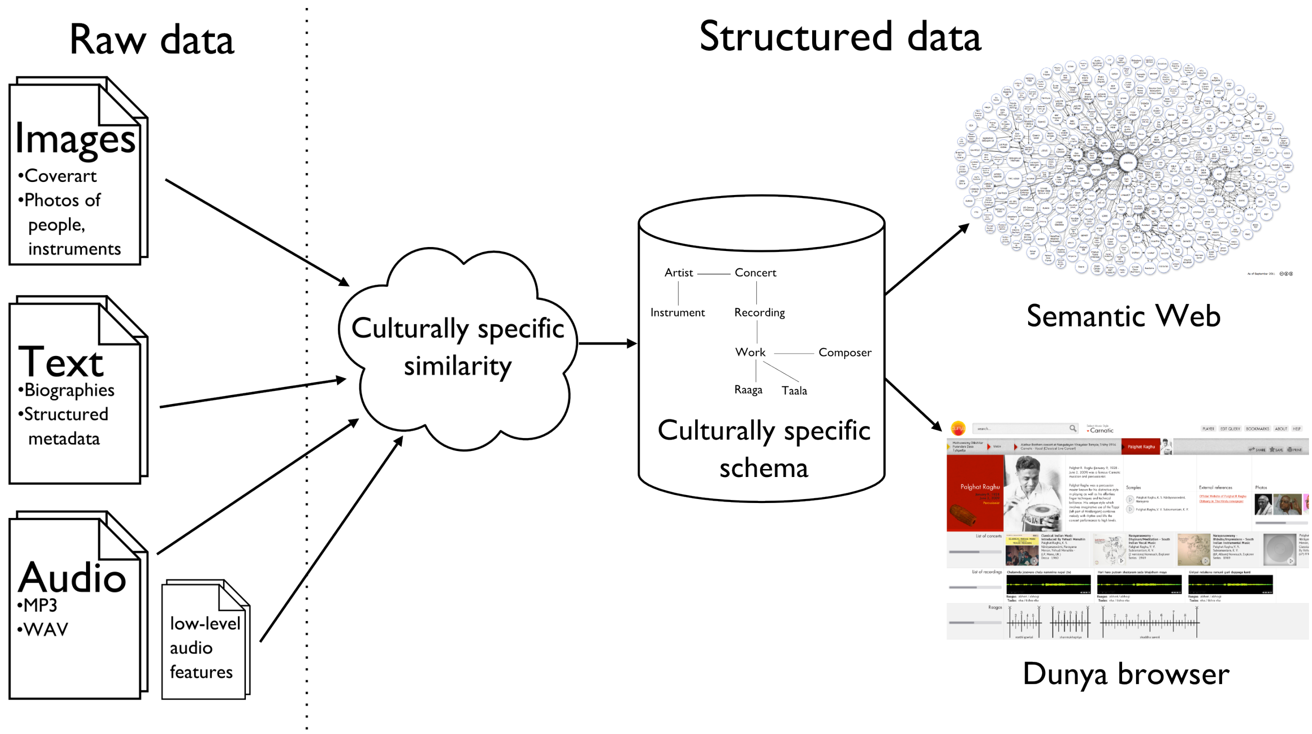

Figure 1. An architecture overview of Dunya. We gather raw data from a number of sources and combine them in a structured database.

for exploring this structured data. The following sections describe these three aspects in more detail.

3. RAW DATA

3.1 Audio recordings

The audio recordings are the core of Dunya. In order to build our audio collection, we bought 300 audio CDs, containing approximately 380 hours of audio and 1900 distinct recordings. We will be extending this collection and hope to reach 500 hours of audio in the next two years. We ripped these audio CDs, and stored the recordings in our file server.

3.2 Editorial metadata

Every audio recording of the collection is accompanied by editorial metadata. Since most audio recordings come from commercial CDs, the editorial metadata comes from the cover or the booklet accompanying the CD. We use MusicBrainz [2] to store and organise this metadata, which includes names of recordings, releases, compositions, composers, artists, and other culture-specific musical concepts. Though Musicbrainz contains metadata for a large number of audio and albums, most of the metadata of the audio recordings that we obtained was not yet in MusicBrainz and so we added it ourselves.

MusicBrainz is a very useful information source for us for a number of reasons. MusicBrainz provides stable identifiers for musical entities, in the form of MusicBrainz Identifiers (MBIDs). These IDs are guaranteed to be unique over all items in the MusicBrainz database. This makes it possible to unambiguously refer to a musician, composition, recording, or released album. For this project we use MBIDs to refer to all entities in our corpus. We see MusicBrainz as a useful tool in many fields of MIR research to explicitly identify datasets used in research. It is able to represent additional relationships in addition to the typically used artist-release-recording model. An example relationship that we preserve in our dataset is the information indicating the people who performed on albums and the instruments that they played. All factual data entered into MusicBrainz becomes freely available under the MusicBrainz project's licenses to be used for any desired use. For a research project such as ours the free availability of the data makes it a useful source for factual information.

3.3 Biographical metadata

Where they exist we use Wikipedia for biographies of artists in our database. When MusicBrainz lists a link to Wikipedia as the artists biography, we use that page. Where such a link doesnt exist we create a link between MusicBrainz and Wikipedia as part of our import process.

3.4 Images

We gather representative images for each item. We get cover art images for commercial CDs from the Cover art archive, [3] a free repository of CD cover art images. We add cover images to the Cover art archive whenever we add new CDs to MusicBrainz. Images of artists and composers

[2] `http://musicbrainz.org`

[3] `http://coverartarchive.org`

are currently obtained from Wikipedia and other community websites related to the culture whose music we are working with. Images of instruments used in the musical culture are taken from Wikipedia and from repositories containing images with a freely available license such as Flickr.

3.5 Audio features

We compute low to mid level audio features from the audio recordings in our music collection and store them with the audio recordings. We are currently focusing on melodic and rhythmic dimensions of the music and so have extracted low level audio features such as perceptual amplitude, onsets, and predominant pitch [12]. For the Indian music collections, we also extract the tonic pitch of the lead performing artist [13]. We use the Essentia library [1], an audio analysis library developed by our research group that includes most of the common low and mid level feature analysis algorithms, to compute these features. We are currently performing research on various culture specific descriptors that will be integrated into Essentia and used in Dunya as they become available. For example we are currently working on intonation analysis [6], motivic analysis [11], and *rāga* characterisation [5]. The audio features are stored in the system using the YAML format. [4]

4. CULTURALLY STRUCTURED DATA

After we have gathered all of the raw data that is part of a given corpus we organise it and store it in a database. The information is stored using a schema that is designed to reflect the culture that we are working with. We are currently working to develop cultural-specific schemas for all music cultures that are part of the CompMusic project. For this first version of the system, we have developed a basic schema for Carnatic music which we discuss in Section 5. In this section we discuss the main types of structured data that we create (on top of the raw data) and display in our corpus browsing application.

4.1 High level audio features

Carnatic music is based on well-established melodic and rhythmic frameworks, *rāga* and *tāla* respectively. It is heterophonic with the most dominant characteristics of the music arising from melody. Each melody is set in a specific *rāga*. The main characteristics of *rāga* are the constituent svaras (musical notes), their ascending and descending progressions, their role, and gamakas (ornamentations) sung/played with them, and characteristic phrases [7]. Several descriptors are extracted from the audio for a comprehensive description of melodies. These include tonic, F0 contours, loudness and timbre of the predominant melodic source. Given a *rāga*, the intonation of each svara is a function of its role and the gamakas sung with it. A compact description of the overall intonation of svaras is extracted using pitch histogram of a given recording [6]. Melodic motives are extracted to retrieve the characteristic phrases of the *rāga* [4, 11]. Intonation description and motives are then used to establish similarity between performances. These, coupled with the pitch contours [1] and the tonic provide a good account of *rāga*.

When characterising the rhythmic structure of Carnatic music the *tāla* cycle is the basis for repetitions of phrases and motives. Tracking the progression of these cycles is essential for a rhythmic description of a piece. Each *tāla* cycle is divided into metrical positions, sometimes unequally spaced. We extract these metrical positions of the *tāla*, which are often called the beats of the *tāla*. Of all the positions, the first metrical position of the cycle, called the sama is the most important. The work towards the complete description of the *tāla* is in progress. Though tempo is not clearly defined in Carnatic music, we presently extract the rhythmic density based on the onsets, which is an indicator of the tempo of the piece.

The characterisation of rhythmic and melodic phrases in Carnatic music and the extraction of these phrases is research in progress. In addition we are also working towards the structural segmentation of the music piece at different time scales such as motives of a phrase, phrases of a piece, and pieces of a concert. These structural segmentation features provide a perspective about the structure of the performance and all these melodic description components are the building blocks for developing melodic similarity measures.

4.1.1 Analysing audio

We have developed an audio processing framework that is used to compute the high level audio features from the audio recordings in our corpus. Because the project is still under active development it is important to be able to reanalyse the audio whenever our analysis algorithm is modified, and to run the analysis on newly added audio. Our audio processing is performed on a cluster of servers. Each server in the cluster has access to a shared NFS disk that contains the entire audio corpus. A central PostgreSQL database keeps track of which audio files have been analysed by the algorithms. We use the Celery task queue system [5] to distribute the analysis tasks to the machines in the cluster. When audio is added to the corpus the cluster master directs a worker machine to analyse an album using a specific algorithm. Each worker performs analysis on its workload and stores the results with the audio files, where they become available to our browser application. By performing the analysis using a cluster of servers we are able to add more servers as our corpus grows. The result of the analysis is stored in a log, allowing an administrator to quickly check if there there was a problem with the analysis of an album.

4.2 Structured editorial metadata

Our system stores structured editorial metadata for all items in the corpus in a PostgreSQL database. This includes information from MusicBrainz as well as information that

[4] http://www.yaml.org

[5] http://www.celeryproject.org

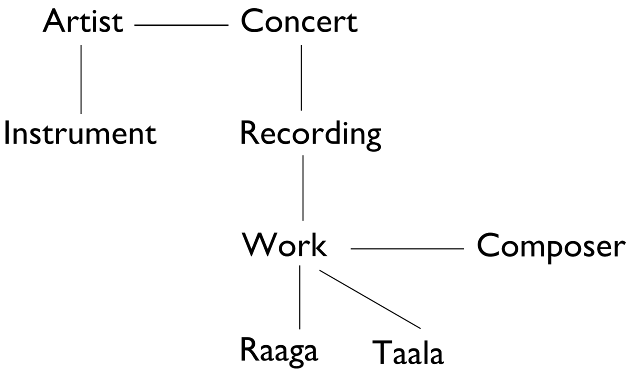

Figure 2. The main entity types in Carnatic music and their relationships.

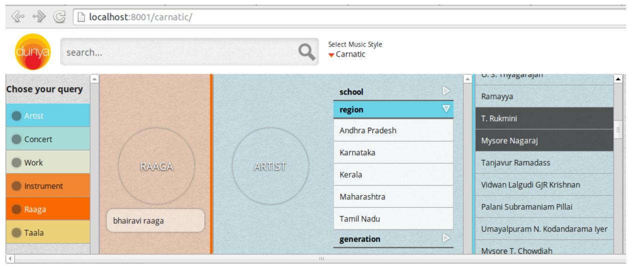

Figure 3. The search navigation interface showing a query for items related to a specific *rāga* and two artists.

we have obtained from other online sources such as Wikipedia or websites that contain information and discussions related to the style of music that we are working with. Some information can be represented as ontologies, explicitly expressing the relationships between concepts used in that culture. We store the RDF ontologies in a graph database allowing us to navigate through related concepts in our database. We initially attempted to access data on demand via APIs provided by each service that provides data for our corpus, however we abandoned this approach for two reasons. Firstly, the time taken to perform queries on all dependent sites in order to display the relevant information on a page would take too long. Secondly, by controlling all of the metadata in our database we are able to create relationships between items in the database that may not exist in the other data sources we are using. We keep references to all of our external sources in order to direct users to them if they wish to find more information. By storing the date at which data was taken from an external source we can periodically refresh the data by checking if the source information has changed since we last accessed it. Because we store all of the relevant information in a local database it is also possible to take the entire collection and access it locally in situations where an Internet connection is not available.

5. BROWSER

The Dunya front-end is a web-based application written in Python using the Django web framework, [6] with an interactive interface written in Javascript. The interface is designed to work on a range of computing devices.

We also plan to make the same data that we show in the browser available in a machine-readable format. To do this we will use the Music Ontology initiative [10] to provide an endpoint that lets people write systems to query the items in our database using publicly known identifiers such as MBIDs.

For the Carnatic music collection we focus on seven main concepts that are important to the culture surrounding the music. The concepts are: artist, composer, work, recording, concert, *rāga*, and *tāla*. We refer to each instance of one of these concepts as *entities*. For every entity in our database we are able to show the information that we have gathered from linked sites as well as suggestions for other entities that are related to the currently viewed one. *rāga* and *tāla* are concepts that are represented as attributes of a work. These works are performed by musicians at concerts. Concerts are a fundamental musical unit in the Carnatic culture and so we use these as a type of entity. Almost all albums of this kind of music are recordings of live concerts. We have developed ontologies for the *rāga* and instrument entity types for Carnatic music.

5.1 Search and discovery

In Dunya we provide two different ways to find music and other items in the collection. The search interface allows a user to type in the name of any entity or concept that we store in the database. As the user types in the search box we suggest recommendations that they might wish to use. We show results of all entities that contain the text that the user has entered. When searching we also consider alternative spellings of entities due to different ways of romanising Indian text to Latin characters. We make use of alternative spellings entered in MusicBrainz, as well as a fuzzy Levenshtein distance-based measurement.

The second search interface is an interactive browser (Figure 3). For each of the core concepts (artist, composer, work, concert, *rāga*, and *tāla*) we show filters that allow users to find entities that match certain criteria. For example, an artist can be filtered by a time period, the musical school that they trained in, or the region that they are from. The data for these filters comes from online sources such as biographies. A user can query with as many concepts as they like. As more queries are added the number of relevant results returned is reduced.

5.2 Entity pages

Once the user selects an entity in the search results we show a page describing it. At the top of the page we show editorial and relationship data. This information comes from the information that is present in Musicbrainz, as well as from other linked information that we have discovered. For example, the page for an artist shows basic biographical information such as their name and birth and death dates. The page includes biographical information if we

[6] `https://www.djangoproject.com`

Figure 4. Visual display of a recording in Dunya. Suggestions for similar recordings are shown underneath the audio player.

have been able to source it with a link to the original entry. Where available we also show an image of the entity. Artist pages show an image of the person, while for concerts we show the cover art of a CD if available. We also show links to other related entities. For example, the page for an artist will show concerts that they have played in and the instrument that they play. The page for a *rāga* shows composers that often wrote compositions using that *rāga* and popular compositions written using it.

One core concept in Dunya is to be able to listen to a relevant sound sample for each entity. In the display of each entity there is a button that when pressed plays a sound fragment that has been automatically chosen for its relevance. From the chosen fragment a user would be able to access the entity of the full recording that the segment comes from. Audio recordings are added to a global playlist that continues playing as the user navigates around the corpus. At any point in time the user can view the playlist, edit it, or go to the recording page of the currently playing track.

5.3 Recording entity

One of the primary entity interfaces that we are developing in Dunya is the view that represents a single audio recording. Most of the technologies developed in this project result in new methods to analyse and understand audio and the recording page is where we visually show the results of the analysis while playing the recording (Figure 4). The recording interface overlays the extracted features on top of a visual representation of the recording. In this version of the system we show two types of low-level audio features and three types of higher level features on the recording page.

We show two different representations of the audio. The upper view shows a time-domain waveform, and the lower one shows a spectral view. These views can be zoomed in to see data in more detail. For performance reasons we pre-render the low-level visualisations at different zoom levels and draw the features on top of them. Between the two views there is a full-length representation of the time-domain signal. This gives an indication of the progress through the track even if the other views are zoomed in. On the time-domain waveform we show extracted rhythmic information. This information indicates the metrical positions of the *tāla* of the recording. On the lower spectrogram image we show the pitch contour of the extracted melody of the recording. To the left of the spectrogram is a histogram of the predominant pitches in the entire recording. We indicate the tonic pitch of the recording with a horizontal line across this image. A playback cursor is shown on both of these visualisations as the recording is being played back. As the recording plays we indicate the current predominant pitch on the histogram to the side of the melody pane.

Below the audio player we show recordings that are similar according to the distance metrics that we have developed. Currently we show recordings that have a similar *tāla* (metrical framework) and *rāga* (melodic framework). The list of recordings is ordered by the similarity of the recordings to the current recording. The first recordings in the list may have the same *rāga* as the current recording, however following recordings may use *rāgas* that are are related based on the ontologies that we have developed.

6. CONCLUSIONS

We have presented Dunya, a music discovery system being developed as part of the CompMusic project designed to support the exploration of a music corpus using concepts relevant to the particular musical culture of the corpus. The application displays the typical information shown in music playback applications, but also includes additional information collected from a variety of sources and information that has been automatically computed from the recorded audio. It stores a complex set of relationships between the items in the database. With this information a user can explore given music collection and discover musically relevant relationships. The architecture of the system and the interface have been designed to support cultural specificity, starting in this paper with Carnatic music. In the near future we will extend the system to the rest of the music collections we are working with and we will continue integrating research results resulting from the CompMusic project to enhance the discovery capabilities of Dunya.

7. ACKNOWLEDGEMENTS

The authors would like to thank Ajay Srinivasamurthy, Gopala K. Koduri, and Sankalp Gulati for contributing to the descriptions of their respective research that is being used in the CompMusic project. Pere Esteve assisted with the design and development of the graphical interface of the Dunya Browser. The CompMusic project is funded by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) / ERC grant agreement 267583

8. REFERENCES

[1] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José Zapata, and Xavier Serra. Essentia: An Audio Analysis Library For Music Information Retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference*, 2013.

[2] Chris Cannam, Christian Landone, Mark Sandler, and Juan Pablo Bello. The sonic visualiser: A visualisation platform for semantic descriptors from musical signals. In *Proceedings of the 7th International Conference on Music Information Retrieval*, 2006.

[3] Masataka Goto, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano. Songle: A web service for active music listening improved by user contributions. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 311–316, 2011.

[4] Vignesh Ishwar, Ashwin Bellur, and Hema A Murthy. Motivic analysis and its relevance to raga identification in carnatic music. In *Proceedings of the 2nd CompMusic Workshop*, 2012.

[5] Gopala K. Koduri, Sankalp Gulati, Preeti Rao, and Xavier Serra. Raga Recognition based on Pitch Distribution Methods. *Journal of New Music Research*, 41(4):337–350, 2012.

[6] Gopala K. Koduri, Joan Serrà, and Xavier Serra. Characterization of intonation in carnatic music by parametrizing pitch histograms. In *Proceedings of the International Society for Music Information Retrieval Conference*, pages 199–204, 2012.

[7] TM Krishna and Vignesh Ishwar. Carnatic music: Svara, gamaka, motif and raga identity. In *Proceedings of the 2nd CompMusic Workshop*, 2012.

[8] François Pachet, Jean-Julien Aucouturier, Amaury La Burthe, Aymeric Zils, and Anthony Beurive. The cuidado music browser: an end-to-end electronic music distribution system. *Multimedia Tools and Applications*, 30(3):331–349, 2006.

[9] Elias Pampalk and Masataka Goto. Musicsun: A new approach to artist recommendation. In *Proceedings of the 8th International Conference on Music Information Retrieval*, pages 101–4, 2007.

[10] Yves Raimond, Samer Abdallah, Mark Sandler, and Frederick Giasson. The music ontology. In *Proceedings of the 8th International Conference on Music Information Retrieval*, pages 417–422, 2007.

[11] Joe Cheri Ross, TP Vinutha, and Preeti Rao. Detecting melodic motifs from audio for hindustani classical music. In *Proceedings of the International Society for Music Information Retrieval Conference*, 2012.

[12] Justin Salamon and Emilia Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech and Language Processing*, 20:1759–1770, 2012.

[13] Justin Salamon, Sankalp Gulati, and Xavier Serra. A multipitch approach to tonic identification in indian classical music. In *Proceedings of the International Society for Music Information Retrieval Conference*, 2012.

[14] Xavier Serra. A Multicultural Approach to Music Information Research. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011.

[15] George Tzanetakis. Musescape: An interactive content-aware music browser. In *Proceedings of the 6th Conference on Digital Audio Effects (DAFX)*, 2003.