**Computing Lab and Data Warehousing & BI Project – Milestone III**

Niti Mishra, Miquel Torrens, Bálint Ván

**Database descrption**

The database created for the project is named omsong. It is build with four interrelated tables that act as data warehouse for the analytics part. You can see the schema attached to this document. The four tables are the following:

1. song\_metadata: this table contains the most part of the information, with the details at a song level: aggregates on the technical features (like tempo, loudness…), and also metadata such as artist, album, year or location. It contains 10,000 observations and has as primary key the song\_id. It also contains the artist\_id to link with other tables. At the moment the table is indexed on these two columns. If the analytics require extra indexes we will add them as we go.
2. artist\_relations: it contains the relations between artists, that is, what are the most usually related artists to each artist. It features what artists are related and in what rank. The artist\_relation\_id is the primary key of the table and it is indexed by this field and also by the song\_id and the artist\_id fields, which are foreign keys song\_metadata to the table.
3. artist\_terms: this table contains the terms (e.g genre) associated to each song referring to the artist. It also features the frequency and the weight of its terms. The term\_artist\_id is the primary key of the table and it is indexed by this field and also by the song\_id and the artist\_id fields, which are foreign keys song\_metadata to the table.
4. song\_tags: it contains the tags associated to each song gathered from Musicbrainz. It also features the frequency of the terms. The song\_tag\_id is the primary key of the table and it is indexed by this field and also by the song\_id and the artist\_id fields, which are foreign keys song\_metadata to the table.

We are still working on adding a fifth additional table that will incorporate the words in the lyrics associated to each song (linked by song\_id). Due to tractability issues with the source files (completely unstructured data), this table is still not running on the MySQL Database.

These tables will eventually be complemented with some summary tables or functions/procedures on the results of the analysis. This second part will be designed depending on the analytics outcomes. This second part of the database will be built once these outcomes are established.

**Schema for the** omsong **database**

