Roadmap

28 SEP - 3 OCT

- Define roadmap
- Create Github repository
- Complete bibliography

4 OCT - 10 OCT

- Text preprocessing
  - Translate non-English texts
  - Define approaches (full plain text, full weighted text, summarization [extractive, abstractive])
  - File generated:
    - *lyrics_plain.txt* - complete lyrics
    - *choruses.txt* - choruses of songs, where possible extracted via metadata, otherwise using extractive summarization algorithm
    - *lyrics_weighted.txt* - complete lyrics, each sentence has a weight: 5 if it belongs to the chorus, 1 otherwise (parameters are chosen arbitrarily and should be discussed)
    - *lyrics_summarized.txt* - each song is represented by a summary, obtained with abstractive summarization techniques (T5 model)

- Audio/vocals preprocessing
  - Consider approaches (full audio, excerpts, song summarization)
  - Extract 30s excerpts, from 30” to 60”
  - Extract 10s excerpts, using an algorithm that tries to identify chorus detecting pattern repetition in frequencies (pychorus)
  - Music-voice separation in 30s and 10s excerpts
  - Choose useful features

11 OCT - 17 OCT

- Prepare input for models:
  - Vocal features using openSmile (eGeMAPSv02): *vocals_full.csv, vocals_30s.csv, vocals_10s.csv*
  - Music features using pyAudioAnalysis (high-level features): *music_full.csv, music_30s.csv, music_10s.csv*
  - Text features using SSWE and sBERT

- Train models with single input
  - Text-based models
  - Audio-based models
  - Voice-based models

18 OCT - 24 OCT

- Evaluate classification performances (genre/emotion)

- Define multimodal approaches (high/mid/low-fusion level and combinations)
- Prepare input for multimodal models
- Build high level fusion multimodal models

- Build low level fusion multimodal models
- Evaluate classification performances (genre/emotion)

25 OCT - 31 OCT

- Global performances evaluation
- Genre-emotion correlation
- Model tuning