

# MusicStats

Luís Dias

Faculdade de Engenharia da Universidade do Porto  
080509094

Luís Gomes

Faculdade de Engenharia da Universidade do Porto  
080509169

21 de Novembro de 2012

## Resumo

O MusicStats consiste na concatenação da informação de dois datasets diferentes, o Last.fm e o MusixMatch. Assim, através do cruzamento destes dois tipos de informação, criámos um novo dataset. Indexando os dados mais importantes, tornando assim a pesquisa menos morosa, conseguimos oferecer aos utilizadores as mais variadas informações acerca das músicas, tais como, quais as palavras mais usadas nas letras das músicas ou quais os temas de canções mais vendidas.

## 1 Introdução

Qual a banda de música que não gostaria de saber, à priori, que hipóteses tem da sua música nova de ser ouvida? Ou de todas as músicas existentes, quais os temas de canções que mais vende? Foi dessa necessidade que nasceu o MusicStats. No fim deste trabalho, pretende-se ter um dataset construído por nós, que servirá para auxiliar a análise musical actual. A partir daí, e com a informação presente, a imaginação é o limite, e será possível conjugar todos os dados das mais ínfimas formas. Saber qual o tipo de música mais usado, qual o tema de música tem mais influência nas músicas, quais as palavras mais usadas ou qual o tema preferido por cada banda aquando da criação das músicas. Para isso, criámos um dataset, que reusulta da concatenação entre o dataset da plataforma online, Last.fm, e o MusixMatch. Se no primeiro existe toda a informação de todas as músicas de um determinado artista, o segundo contempla todas as palavras pertencentes a uma letra de uma música. Assim sendo, o novo dataset por nós criado, para além das informações da música, contém to-

das as palavras presentes na sua letra.

Para realizar este trabalho foram consultadas variadas referências bibliográficas relevantes, tais como

- <http://labrosa.ee.columbia.edu/millionsong/lastfm> Plataforma que contém toda a informação relevante do dataset do Last.FM
- <http://www.lastfm.com.br/api> Plataforma que contém toda a informação relevante da API do Last.FM
- <http://labrosa.ee.columbia.edu/millionsong/musixmatch> Plataforma que contém toda a informação relevante do dataset do MusixMatch
- <https://developer.musixmatch.com/documentation> Plataforma que contém toda a informação relevante da API do MusixMatch

## 2 Ferramenta a usar

Para este trabalho, escolhemos a ferramenta *open-source* Lucene. Com ela, podemos indexar o conjunto de dados que pretendemos, aumentando assim a velocidade das consultas. Para além disso, é uma ferramenta com imensa documentação online, o que permite uma rápida e fácil aprendizagem dos processos internos da ferramenta.

## 3 Geração dos Índices

Como pretendemos dar resposta às interrogações dos utilizadores, teríamos obrigatoriamente

que gerar índices para aumentar a rapidez de resposta. No entanto, não faria sentido indexarmos todos os campos do dataset, em primeiro lugar porque seria um processo bastante moroso e, em segundo lugar, porque não faria sentido indexar, por exemplo, campos relativos ao ID da música já que não iria haver utilizadores a fazer interrogações por esse campo. Assim, estimámos quais os campos de pesquisa que mais vezes vão ser utilizados, e foram esses os escolhidos para a indexação. Desta forma, indexámos os campos *lyrics* - que tem todas as palavras da letra da música - e *artist*, que tem o nome do intérprete dessa música. Para isso, e tal como referido na secção anterior, utilizámos a biblioteca Lucene, que contempla funções de criação e gestão de índices, assim como auxiliares de pesquisa de informação.

## 4 Demonstração do processo de interrogação

Como referido anteriormente, após o dataset estar devidamente construído, a sua consulta deve responder a uma variedade de perguntas, entre as quais podemos destacar:

- Quais as palavras mais usadas em todas as canções;
- Qual o género de músicas mais usado;
- Qual o tema mais utilizado na composição de músicas;
- Qual a Banda que mais músicas tem sobre determinado tema;
- Quais as músicas que falam sobre determinado tema.

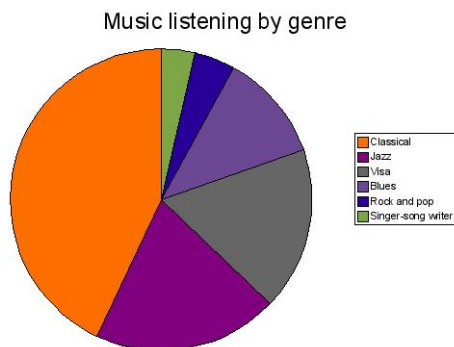


Figura 1: Há géneros que são mais ouvidos que outros, mas terá isso alguma ligação com o tema?

## 5 Conclusões

Este sistema servirá para o correcto estudo e análise dos temas utilizados pelos compositores no acto de criação das próprias músicas. Nos tempos que correm, deve cada vez mais ser prestada a devida atenção ao meio envolvente, ao sucesso alheio. Esta aplicação vem, precisamente, ao encontro dessa necessidade na medida em que estuda todos os elementos existentes no mercado, fazendo uma espécie de análise de êxito ou de fracasso. Analisa todas as canções para sabermos o que é mais utilizado, para, ou serem seguidas as mesmas directrizes, numa tentativa de sucesso seguro, ou tentar criar algo novo e inovador.

## Referências

- [1] Million Song Database. The Last.fm Dataset, URL: <http://labrosa.ee.columbia.edu/millionsong/lastfm>, Outubro 2012.
- [2] Last.fm. API - Last.fm, URL: [www.last.fm/api](http://www.last.fm/api), Outubro 2012.
- [3] Million Song Database. The musiXmatch Dataset, URL: <http://labrosa.ee.columbia.edu/millionsong/musixmatch>, Outubro 2012.
- [4] MusiXmatch. Documentation, URL: <https://developer.musixmatch.com/documentation>, Outubro 2012.