



Universidade do Estado da Bahia
Departamento de Ciências Exatas e da Terra
Colegiado de Sistemas de Informação

MUSID7: Um Sistema para Recomendação de Músicas Integrando ID3 e MPEG-7

Marcelo Tonete de Aragão

Salvador-BA

2014

Marcelo Tonete de Aragão

MUSID7: Um Sistema para Recomendação de Músicas Integrando ID3 e MPEG-7

Monografia submetida ao Colegiado de Sistemas de Informação da Universidade do Estado da Bahia como parte dos requisitos necessários para obtenção do grau de Bacharel em Sistemas de Informação.

Área de Concentração: Sistemas de Informação
Linhas de Pesquisa: Recomendação de Músicas

Alexandre Rafael Lenz
(Orientador)

Salvador-BA
2014

Resumo

O uso crescente de arquivos multimídia tem trazido um problema referente ao gerenciamento eficiente desses recursos devido ao grande volume disponibilizado na principal fonte de arquivos audiovisuais, a internet. O gerenciamento de conteúdo multimídia se torna necessário para levar ao usuário o retorno desejado de informações audiovisuais de seu interesse. As Etiquetas ID3 vem como uma alternativa para a organização de arquivos de música ao usar metadados para agrupar músicas com uma determinada característica semelhante. Por outro lado, o MPEG-7 define um conjunto padronizado de ferramentas audiovisuais que criam descrições de um arquivo multimídia. Essas descrições podem formar a base para aplicações de recuperação de arquivos audiovisuais ao fornecer as características necessárias para o acesso a um conteúdo multimídia contido em uma base de dados. Neste projeto é realizada a implementação de um sistema de recomendação de músicas baseado em gênero musical que utiliza dois conjuntos de metadados: as etiquetas ID3 e os descritores do MPEG-7. Essa implementação é uma proposta para aumentar a eficiência na recomendação de músicas em relação a utilização individual de um dos conjunto de metadados.

Palavras-chave: Recomendação de músicas, MPEG-7, Etiquetas ID3, descritores multimídia, metadados.

Abstract

The growing use of multimedia files has brought a problem regarding the efficient resource management due to the large volume of data available in the main source of audiovisual archives, the Internet. The multimedia content management system becomes necessary to bring to the user the desired return of audiovisual information. ID3 tags come as an alternative for organizing music files using metadata to group songs by similar determined feature. On the other hand, the MPEG-7 defines a standard set of audiovisual tools that creates descriptions for a multimedia file. Those descriptions can form the basis for audiovisual archives recovery applications by providing the necessary features for the access to a multimedia content contained in a database. In this project, the implementation of a music recommendation system based on musical genre is performed using two metadata sets: the ID3 tags and descriptors of MPEG-7. This implementation is a proposal to increase the song recommendation efficiency compared to individual use of one of the metadata sets.

Keywords: Songs Recommendation, MPEG-7, ID3 tags, multimedia descriptors, metadata.

Agradecimentos

Agradeço a DEUS pela força e pela família que me deu a qual me guiou de forma plena para todas as conquistas obtidas em minha vida.

Agradeço aos meus pais, Maurício Antonio Lustosa de Aragão (In Memoriam) e Leoní Santos Tonete que me apoiaram em tudo e me deram todos os meios de obter mais essa vitória apesar de todos os sacrifícios e surpresas que a vida nos trouxe. Aos meus irmãos Maurício Tonete de Aragão e Marcus Tonete de Aragão aos quais eu dedico mais essa vitória.

Agradeço a minha esposa Paula Teixeira de Aguiar por estar ao meu lado desde o início dessa jornada me apoiando em tudo, principalmente nos momentos mais difíceis.

Agradeço aos amigos Raylan Souza, Kal Lenon, Leonardo Campos, Daniel Amaral, Thomas Magnum, Felipe Zacarias pela companhia ao longo desse difícil caminho onde juntos conseguimos superar diversos obstáculos.

Agradeço aos meus chefes, comandantes e amigos Major PM Washington Eurico Silva Costa e Capitão PM Cristiano Jose de Oliveira Paraíso que me deram todo o apoio necessário para seguir nessa jornada de mais uma graduação.

Agradeço a meu orientador, Alexandre Lenz, por ter apoiado meu projeto e me orientado em tudo que precisei para a conclusão.

"A lot of times, people don't know what they want until you show it to them."

Steve Jobs

Sumário

| | | |
|-------|--|----|
| 1 | Introdução | 12 |
| 2 | Sistemas de Recomendação de Conteúdo | 16 |
| 2.1 | Contextualização | 16 |
| 2.2 | Recomendação Baseada em Conteúdo | 18 |
| 2.3 | Filtragem Colaborativa | 18 |
| 2.4 | Sistemas Híbridos | 19 |
| 3 | Metadados | 20 |
| 3.1 | Etiquetas ID3 (<i>Identify a MP3</i>) | 20 |
| 3.2 | O Padrão MPEG-7 | 22 |
| 3.2.1 | Conceito | 22 |
| 3.2.2 | Elementos do MPEG-7 | 25 |
| 3.2.3 | Componentes do MPEG-7 | 28 |
| 4 | Trabalhos Relacionados | 31 |
| 4.1 | Classificação de Gêneros Musicais de Sinais de Áudio | 31 |
| 4.2 | Uma Pesquisa de Sistemas de Recomendação de Músicas e Perspectivas Futuras | 32 |
| 4.3 | Nextone Player: Um Sistema de Recomendação de Músicas Baseado no Comportamento do Usuário | 33 |
| 4.4 | Classificação de Áudio Baseada em Extração de Características e Redes Neurais | 35 |
| 5 | Arquitetura do Sistema Para Uso Individual de Metadados | 38 |
| 5.1 | Arquitetura | 39 |
| 5.2 | Implementação | 40 |
| 5.2.1 | Método de Extração das Etiquetas ID3 | 40 |

| | | |
|-------|---|----|
| 5.2.2 | Método de Extração dos Descritores MPEG-7 | 42 |
| 5.2.3 | Módulo de Banco de Dados do Sistema | 47 |
| 5.2.4 | Módulo de Recomendação | 51 |
| 5.3 | Resultados | 54 |
| 5.3.1 | Análises Experimentais | 57 |
| 6 | MUSID7: Um Sistema para Recomendação de Músicas Integrando ID3 e MPEG-7 | 61 |
| 6.1 | Arquitetura | 61 |
| 6.2 | Implementação | 63 |
| 6.3 | Resultados | 65 |
| 7 | Conclusões e Trabalhos Futuros | 67 |
| 7.1 | Conclusões | 67 |
| 7.2 | Trabalhos Futuros | 70 |
| A | Apêndice A | 75 |
| B | Apêndice B | 78 |

Lista de Figuras

| | | |
|-----|---|----|
| 2.1 | Recomendação de Conteúdo | 17 |
| 3.1 | Uso do ID3 | 21 |
| 3.2 | Abstrações das Versões do MPEG. | 24 |
| 3.3 | Vetor de Características do Descritor <i>AudioPower</i> | 26 |
| 3.4 | Plotagem do Descritor <i>AudioPower</i> | 27 |
| 3.5 | Modelo de Esquema de Descrição do MPEG-7. | 28 |
| 3.6 | Exemplo de Esquema de Descrição do MPEG-7. | 29 |
| 4.1 | Modelo de Árvore para Classificação Automática de Gêneros Musicais | 32 |
| 4.2 | Aproximação Linear do Envelope de Um Sinal de Áudio | 36 |
| 5.1 | Arquitetura do Sistema. | 40 |
| 5.2 | Representação dos Descritores <i>AudioPower</i> e <i>Waveform</i> | 45 |
| 5.3 | Representação dos Descritores Espectrais. | 46 |
| 5.4 | Fórmula do Cálculo da Distância Euclidiana. | 48 |
| 5.5 | Limitação Individual das Etiquetas ID3 | 55 |
| 5.6 | Gráfico das Médias dos Tempos de Recomendação | 58 |
| 5.7 | Gráfico das Médias de Porcentagem das Concordâncias nas Recomendações | 58 |
| 5.8 | Exemplo de Hierarquia de Gêneros Musicais. | 59 |
| 5.9 | Gráfico de Médias de Memória Utilizada nas Recomendações | 60 |
| 6.1 | Arquitetura do MUSID7. | 62 |
| 6.2 | Representação da Recomendação do MUSID7. | 63 |
| 6.3 | Média do Grau de Concordância de Gêneros Musicais do MUSID7 | 66 |

Lista de Tabelas

| | |
|--|----|
| 3.1 Limite de Informações da versão ID3v1 | 22 |
| A.1 Relação de Gêneros Extraídos das Etiquetas ID3 | 77 |
| B.1 Adaptações dos Trabalhos Relacionados | 79 |

Lista de Algoritmos

| | |
|---|----|
| 5.1 Algoritmo da Classe Musica | 42 |
| 5.2 Algoritmo da Extração das Etiquetas ID3 | 43 |
| 5.3 Algoritmo do Módulo de Banco de Dados do ID3 | 48 |
| 5.4 Algoritmo do Módulo de Banco de Dados do MPEG-7 | 50 |
| 5.5 Algoritmo do Módulo de Recomendação do ID3 | 52 |
| 5.6 Algoritmo do Módulo de Recomendação do MPEG-7 | 53 |
| 6.1 Algoritmo do MUSID7 | 64 |

Lista de Abreviaturas e Siglas

| | |
|---------|--|
| MPEG | <i>Moving Picture Experts Group</i> |
| XML | <i>eXtreme Markup Languages</i> |
| CESTA | <i>Coletânia de Estidades de Suporte ao uso de Tecnologia na Aprendizagem</i> |
| UFRGS | <i>Universidade Federal do Rio Grande do Sul</i> |
| CUIDADO | <i>Content Unified Interfaces and Descriptors for Audio Databases available Online</i> |
| ID3 | <i>IDentify a MP3</i> |
| ISO | <i>International Organization for Standardization</i> |
| IEC | <i>International Electrotechnical Commission</i> |
| DDL | <i>Linguagem de Definição de Descrição</i> |
| D's | <i>Descritores</i> |
| DS | <i>Esquemas de Descrição</i> |
| MDS | <i>Multimedia Description Schemes</i> |
| MFCC | <i>Mel-Frequency Cepstral Coefficients</i> |

Capítulo 1

Introdução

O uso de equipamentos que reproduzem arquivos multimídia tem aumentado muito nos últimos anos, principalmente com o surgimento de novas tecnologias que buscam satisfazer as necessidades de cada usuário.

A principal fonte de arquivos audiovisuais, a internet, se popularizou com a banda larga, gerando um rápido crescimento de informações digitais. Apesar do fato de que os usuários têm cada vez mais acesso a esses recursos, identificá-los e gerenciá-los de forma eficiente é cada vez mais difícil, por causa do grande volume disponibilizado.

Além disso, o usuário não dispõe de mecanismos de busca eficientes que lhe apresente um retorno com informações audiovisuais satisfatórias que descrevam o objeto procurado. Como por exemplo, o retorno de imagens ou logomarcas com traços semelhantes às linhas desenhadas pelo usuário em um mecanismo de busca.

Os metadados surgem nesse contexto como uma importante ferramenta utilizada para facilitar a recuperação de dados. Os metadados são dados sobre dados, ou seja, são informações sobre um determinado conteúdo e são utilizados para facilitar o entendimento, o uso e o gerenciamento de dados (WIRTHMANN, 2003). Nesse projeto foram utilizados dois conjuntos de metadados: As Etiquetas ID3 e os descritores do padrão MPEG-7.

As Etiquetas ID3 são metadados que vem no arquivo de áudio junto com a própria música a qual referenciam. As Etiquetas são utilizadas nesse projeto como fontes de informações para processos de busca e recuperação de músicas em uma base de dados, pois elas permitem

armazenar diversas informações sobre a música na qual estão inseridas como, por exemplo, título e gênero musical.

Atualmente existem diversas comunidades baseadas em metadados que utilizam as Etiquetas ID3 para realizar processos de recomendação de músicas. O *MusicBrainz* e o *Last.fm* são exemplos dessas comunidades.

O MPEG-7 é um padrão ISO/IEC desenvolvido pela MPEG (*Moving Picture Experts Group*). Também chamado de *Multimedia Content Description Interface*, o MPEG-7 é um padrão para descrever as características de conteúdos multimídia para os usuários pesquisarem e recuperarem estes conteúdos de forma similar ao que ocorre nos mecanismos de buscas na internet (DALLACOSTA; SOUZA; TAROUCO, 2004).

Através de um estudo realizado por Kosch (2002) foi mostrado que as descrições de arquivos audiovisuais dentro do padrão MPEG-7 têm dado o suporte para sistemas de banco de dados multimídia permitindo aos arquivos serem indexados e localizados através de um sistema de busca podendo prover interoperabilidade para o banco de dados em um ambiente distribuído.

O uso do padrão MPEG-7 pode trazer benefícios para diversas áreas, tais como: educação, jornalismo, informação turística, entretenimento, serviços de investigações, sistemas de informações geográficas, arquitetura, aplicações sociais, dentre outras (NACK; LINDSAY, 1999).

Esses dois conjuntos de metadados usados individualmente apresentam limitações que afetam a eficiência do processo de recomendação. Essas limitações podem ser observadas em retornos de músicas de gêneros musicais contrários à preferência do usuário, ou por causa de informações equivocadas presentes nas etiquetas ID3 ou por músicas de diferentes gêneros musicais apresentarem características semelhantes dentro do padrão MPEG-7, por exemplo. Se não houver um método de recomendação eficiente, a volumosa oferta de conteúdo multimídia causa confusão aos usuários e distorções nas classificações das músicas.

Com um recurso eficiente, a recomendação de músicas baseada em gêneros musicais pode proporcionar um ganho efetivo de tempo na filtragem de conteúdo musical fornecendo uma classificação mais fiel aos reais interesses do usuário.

Nesse contexto, esse projeto tem o objetivo de implementar um sistema de recomendação de músicas baseado em gêneros musicais que utilize dois conjuntos de metadados de forma complementar, as etiquetas ID3 e os descritores do padrão MPEG-7. Esse sistema foi batizado de MUSID7 e é validado ao apresentar um possível aumento da eficiência no processo de recomendação de músicas em relação ao uso individual de cada um desses conjuntos de metadados supracitados.

Dessa forma, o usuário é capaz de identificar e até gerenciar as suas músicas através de características semelhantes entre elas. Para alcançar o objetivo proposto nesse projeto, foram realizadas seis etapas.

Na primeira etapa, foi criada uma base de dados com mil títulos musicais a serem disponibilizados para reprodução pelo usuário. Na segunda etapa, ocorreu a extração das informações das Etiquetas ID3 e das características dos descritores do MPEG-7 de todas as músicas que compõe a base de dados da primeira etapa. Na terceira etapa ocorreu a implementação dos módulos de banco de dados e de recomendação mais a geração de duas listas de recomendação de músicas, sendo uma para cada conjunto de metadados usados individualmente.

Na quarta etapa foi realizada uma avaliação em cada lista de recomendação para obter o grau de eficiência do uso de cada conjunto de metadados. A quinta etapa contemplou a implementação de um sistema de recomendação de músicas que utiliza os dois conjuntos de metadados de forma integrada, apresentando ao usuário uma nova lista de recomendação. Na sexta etapa foi avaliada a eficiência do sistema implementado na etapa anterior para a validação.

Esse projeto foi dividido em sete capítulos. O Capítulo 1 apresentou a introdução acerca do tema e da problemática trabalhada. No Capítulo 2 foi apresentado os conceitos e classificação de sistemas de recomendação de conteúdo. No Capítulo 3 foram apresentados os conceitos de metadados, etiquetas ID3 e MPEG-7. No Capítulo 4 foram apresentados os trabalhos relacionados a esse projeto.

No Capítulo 5 foi apresentada a arquitetura do sistema de recomendação para uso individual de metadados, os métodos de extração dos metadados, os algoritmos dos módulos de banco

de dados e recomendação, e os resultados experimentais. No Capítulo 6 foi apresentado o algoritmo do MUSID7 e a avaliação dos resultados de suas recomendações. E no Capítulo 7 foram apresentadas as conclusões e as propostas para trabalhos futuros.

Capítulo 2

Sistemas de Recomendação de Conteúdo

Neste capítulo são expostos alguns conceitos presentes no contexto do projeto. Na seção 2.1 são contextualizados os conceitos referentes a sistemas de recomendação. Na seção 2.2 é conceituada a recomendação baseada em conteúdo. Na seção 2.3 e na seção 2.4 são apresentados a filtragem colaborativa e sistemas híbridos, respectivamente.

2.1 Contextualização

A grande quantidade de informações disponibilizadas na internet permite ao usuário acessar uma grande diversidade de opções. Na maioria das vezes, o usuário não possui nenhum parâmetro de seleção de conteúdo, tendo poucas chances de realizar uma escolha relevante de informação em um tempo hábil. A recomendação de conteúdo serve para direcionar o usuário a uma informação que seja adequada às suas preferências.

Sistemas de recomendação mudaram a forma como as pessoas encontram produtos, informações, e até mesmo outras pessoas. Eles estudam padrões de comportamento para saber o que alguém vai preferir entre uma coleção de coisas que ele nunca experimentou.

Os sistemas de recomendação auxiliam no aumento da capacidade e eficácia do processo de seleção de uma informação, pois realizam a combinação adequada entre as expectativas dos usuários e os produtos, serviços e pessoas a serem recomendados aos mesmos, sendo esse

processo de combinação um dos grandes desafios desses sistemas (CAZELLA; NUNES; REATEGUI, 2005).

A Figura 2.1 mostra uma combinação entre as preferências do usuário e dados complementares para gerar um sistema de recomendação. Essas preferências podem ser coletadas de forma explícita, implícita ou por inferência.

Na forma explícita, essas informações são coletadas através do preenchimento de um formulário por parte do usuário onde o mesmo informa suas preferências. Na forma implícita, as informações são coletadas através do monitoramento do comportamento do usuário durante a navegação em sites, como páginas visitadas, páginas consultadas, etc. A coleta de dados por inferência consiste em descobrir o perfil do usuário com base no comportamento de outros usuários com padrão de comportamento similar ao seu (BARCELLOS et al., 2007).



Figura 2.1: Recomendação de Conteúdo
Fonte: Adaptada de Souza (2014).

A partir de como a recomendação é feita, os sistemas de recomendação se classificam

basicamente em três categorias: recomendação baseada em conteúdo, filtragem colaborativa e sistemas híbridos.

2.2 Recomendação Baseada em Conteúdo

A recomendação baseada em conteúdo usa as preferências do usuário utilizadas no passado para recomendar produtos semelhantes. Essa recomendação é feita utilizando tags descritoras de itens onde um item é recomendado se ele tiver características semelhantes a outros itens descritos.

Nessa abordagem de recomendação, há vários itens candidatos que são comparados a itens previamente avaliados pelo usuário e os itens de melhores correspondências são recomendados. Em um cenário de recomendação de músicas, por exemplo, um usuário que gosta de Caetano Veloso teria recomendação de outros artistas do gênero MPB.

Recomendações baseadas em conteúdo são limitadas pelos recursos que estão explicitamente associados com os objetos que estes sistemas recomendam (ADOMAVICIUS; TUZHILIN, 2005). Dessa forma, para ter um conjunto suficiente de recursos, o conteúdo deve também estar em um formato que possa ser analisado automaticamente por um computador ou os recursos devem estar atribuídos manualmente aos itens analisados.

Dentro desse contexto da recomendação baseada em conteúdo, esse projeto foi desenvolvido utilizando os metadados como conteúdo para filtragem e recomendação de músicas. Esses metadados serão extraídos e estarão em um formato que o algoritmo de recomendação possa, automaticamente, analisá-los.

2.3 Filtragem Colaborativa

Ao contrário da abordagem de recomendação baseada em conteúdo, essa técnica consiste em recomendar conteúdo a um usuário tentando prever itens de suas preferências com base em itens avaliados previamente por outros usuários.

O Grundy foi o primeiro sistema de recomendação que propôs usar os estereótipos como um mecanismo para a construção de modelos de usuários. Esse sistema tinha como base uma quantidade limitada de informações sobre cada usuário de forma individual. Usando estereótipos, o sistema Grundy construiria modelos de usuários individuais para usá-los na recomendação de livros relevantes para cada um deles (ADOMAVICIUS; TUZHILIN, 2005).

Os usuários, que fazem as avaliações usadas como parâmetro para recomendar um conteúdo a um outro usuário, devem fazer parte de um mesmo círculo social. Dessa forma, a vizinhança do usuário é analisada e então se chega à conclusão que se um usuário gostou de gêneros musicais como MPB e ROCK, por exemplo, um outro usuário que gostou de MPB, e que pertence ao mesmo círculo social do primeiro, também pode gostar de ROCK. Porém esse sistema de recomendação necessita de um grande número de informações sobre o usuário e sua vizinhança para funcionar corretamente.

2.4 Sistemas Híbridos

Um sistema híbrido de recomendação de conteúdo utiliza técnicas tanto da recomendação baseada em conteúdo quanto da filtragem colaborativa, evitando as limitações existentes em ambos os métodos. Alguns estudos mostraram que o uso de métodos híbridos na recomendação de conteúdo podem fornecer recomendações mais precisas quando comparadas as recomendações de técnicas puras.

O sistema de recomendação usado pelo *Netflix Inc.* é um exemplo de sistema híbrido. Nele há uma análise dos hábitos de observação e busca dos usuários semelhantes, ou seja, a filtragem colaborativa, e também a filtragem baseada em conteúdo através da recomendação de filmes que possuem características semelhantes àqueles assistidos anteriormente pelo usuário.

Sistemas híbridos de recomendação de conteúdo também podem ser melhorados por meio de técnicas baseadas no conhecimento, como o raciocínio baseado em casos, a fim de melhorar a precisão e recomendação para resolver algumas das limitações dos sistemas de recomendação tradicionais.

Capítulo 3

Metadados

Os metadados são dados que descrevem outros dados. O uso de metadados torna-se uma alternativa eficiente para auxiliar na manipulação de objetos multimídia (NASCIMENTO, 2008).

Dentro desse contexto, esse projeto traz dois conjuntos de metadados que serão analisados em um processo de recomendação de músicas: as Etiquetas ID3 e os descritores do MPEG-7.

Neste capítulo são apresentados os conceitos desses dois conjuntos de metadados. Na seção 3.1 são apresentadas as etiquetas ID3 e na seção 3.2 o padrão MPEG-7 é descrito e conceituado.

3.1 Etiquetas ID3 (*Identify a MP3*)

Em 1996, um pesquisador chamado Eric Kemp conseguiu inserir dados adicionais, textos relacionados e/ou informações gráficas, em arquivos de áudio possibilitando que arquivos MP3 pudessem receber informações além da própria música. Com o uso frequente desse formato de áudio, tornou-se padrão o uso de etiquetas ID3 nos arquivos MP3 (NILSSON; MAHONEY; SUNDSTROM, 2014).

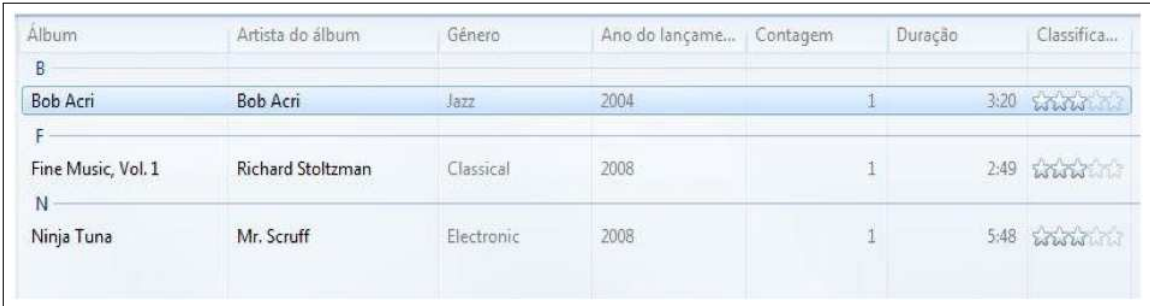
Atualmente não é mais necessário alterar os nomes das músicas para se ter uma ordem

de reprodução dentro das preferências do usuário. As aplicações não mais organizam ou categorizam as músicas por ordem alfabética somente, passando a organizar os arquivos com características semelhantes de cantor, álbum, ano, gênero, etc (JUNIOR, 2014).

A utilização das etiquetas alcançaram todos os principais reprodutores de arquivos MP3 e alguns projetos de usos criativos dessas etiquetas estão sendo desenvolvidos. A reprodução de músicas favoritas, de um determinado usuário, com base em classificações utilizando a frequência de reprodução de músicas é um dos projetos em desenvolvimentos utilizando as Etiquetas ID3 (NILSSON; MAHONEY; SUNDSTROM, 2014).

As etiquetas ID3 são metadados que estão presentes nos arquivos MP3 e que classificam a própria música na qual estão contidos. Todos os dados sobre uma música estão dentro da própria música.

A Figura 3.1 mostra o uso das etiquetas ID3 em um programa de reprodução de áudio, apresentando a organização da lista de reprodução por ordem alfabética baseada no nome do álbum.



| Álbum | Artista do álbum | Gênero | Ano do lançame... | Contagem | Duração | Classifica... |
|--------------------|-------------------|------------|-------------------|----------|---------|---------------|
| B | | | | | | |
| Bob Acri | Bob Acri | Jazz | 2004 | 1 | 3:20 | ☆☆☆☆☆ |
| F | | | | | | |
| Fine Music, Vol. 1 | Richard Stoltzman | Classical | 2008 | 1 | 2:49 | ☆☆☆☆☆ |
| N | | | | | | |
| Ninja Tuna | Mr. Scruff | Electronic | 2008 | 1 | 5:48 | ☆☆☆☆☆ |

Figura 3.1: Uso do ID3

Fonte: Adaptada de Junior (2014).

Durante o desenvolvimento do MP3, as etiquetas ID3 mudaram bastante. Hoje existem duas versões, a ID3v1 e a ID3v2. A diferença principal está na capacidade de armazenamento das informações. A versão ID3v1 contém somente informações sobre o título, artista, álbum, comentário, gênero e ano, sendo que as quatro primeiras informações possuem um limite de 30 caracteres totalizando 128 bytes na estrutura dessa versão.

A Tabela 3.1 mostra a quantidade de informações trazidas na versão ID3v1 e seus respectivos limites de caracteres. Se houver a soma de caracteres de cada campo haverá um resultado

de 125 bytes. Os outros 03 bytes são encontrados no início das etiquetas, antes do título da música.

| Informações | Limite de Caracteres |
|--------------------|-----------------------------|
| Título | 30 Caracteres |
| Artista | 30 Caracteres |
| Álbum | 30 Caracteres |
| Ano | 04 Caracteres |
| Comentário | 30 Caracteres |
| Gênero | 01 byte |

Tabela 3.1: Limite de Informações da versão ID3v1
Fonte: Adaptada de Nilsson, Mahoney e Sundstrom (2014).

Já na versão ID3v2, além das informações trazidas pela ID3v1, outras informações podem estar presentes, como compositor, url, capa do álbum, letras, fotos, etc. E há também um limite superior de caracteres para descrever essas informações. Assim, ao invés de possuir 128 bytes como a ID3v1, a ID3v2 possui 256MB de capacidade sendo que para cada quadro de informação há um limite de 16MB.

Outra diferença entre as versões é encontrada na localização das etiquetas nos arquivos de áudio. A versão ID3v1 contempla a localização das etiquetas no final da música propriamente dita, enquanto que a versão ID3v2 traz essas informações no início deixando a música mais adequada para serviços de streaming.

3.2 O Padrão MPEG-7

3.2.1 Conceito

A partir do início dos anos 90, com o advento do vídeo e áudio digital, as pesquisas em material audiovisual tornou-se importante.

Surgiram sistemas de recuperação de áudio e vídeo baseados na análise do conteúdo desses materiais audiovisuais sendo possível descrever um áudio por sua distribuição de energia

espectral ou por sua frequência fundamental. Assim, foi possível realizar comparações entre áudios utilizando essas descrições sonoras (KIM; MOREAU; SIKORA, 2006).

O MPEG-7 é um padrão de descrição de conteúdos multimídia desenvolvido pelo MPEG (*Motion Picture Experts Group*) e regulamentado pela norma ISO/IEC 15938, *Multimedia Content Description Interface*. Esse padrão provê um rico conjunto de ferramentas de representação de conteúdo audiovisual para propósitos de busca e gerenciamento de objetos multimídia.

Dessa forma, o MPEG-7 fornece interoperabilidade entre sistemas e aplicativos utilizados na geração, gestão, distribuição e consumo de descrições de conteúdo audiovisual. Tais descrições de mídia ajudam os usuários ou aplicativos a identificar, recuperar ou filtrar informação audiovisual. Tanto os usuários humanos quanto os sistemas automáticos que processam as informações audiovisuais estão dentro do escopo do MPEG-7 (MONTEIRO, 2007).

As descrições do MPEG-7 não dependem da forma que o conteúdo descrito é codificado ou armazenado, sendo possível criar uma descrição MPEG-7 de áudio analógico no mesmo modo de um áudio digitalizado.

A Figura 3.2 mostra o grau de abstração das versões do MPEG, sendo o MPEG-1 e MPEG-2 para codificação de áudio e vídeo, o MPEG-4 para compressão de dados digitais de áudio e vídeo além de ser possível a sua implementação com orientação a objetos e o MPEG-7 de forma semântica, como interface de descrição de conteúdo multimídia.

A fim de lidar com uma ampla gama de aplicações, o padrão MPEG-7 fornece elementos normativos como Descritores (D), Esquemas de Descrição (DS), e Linguagem de Definição de Descrição (DDL) sendo eles os três principais elementos que formam o MPEG-7. Além disso, fornece também especificações relativas a questões de sistemas de níveis de descrição, podendo ser especificações para descrições em baixo nível ou especificações para alto nível.

O MPEG-7 aborda uma grande variedade de arquivos de mídia sendo possível a criação de descrição para arquivos de imagens como fotos, gráficos ou modelos 3D, áudio, vídeos ou ainda a combinação destes tipos de arquivos usados frequentemente em apresentações multimídias (HUNTER, 2005).

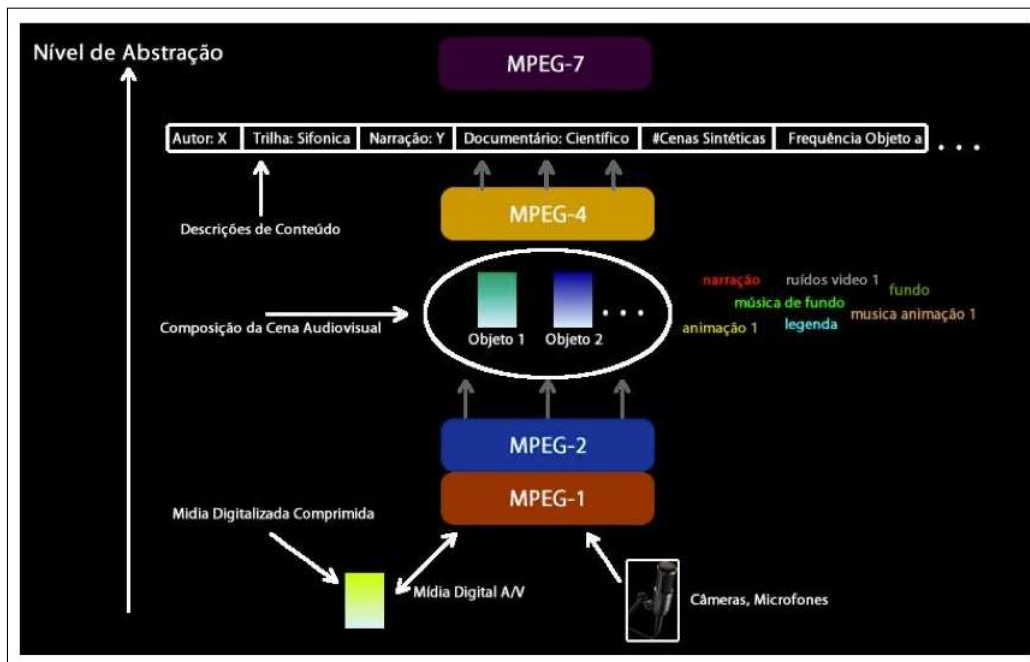


Figura 3.2: Abstrações das Versões do MPEG.
Fonte: Adaptada de Silva e Galindo (2007).

O MPEG-7 não se destina a nenhuma aplicação em particular, não abrangendo a área da extração automática de descritores, deixando os produtores de softwares construir suas próprias ferramentas e aumentar a competitividade entre os seus produtos. O MPEG-7 utiliza o XML como linguagem descritiva, não sendo necessário que a descrição em XML esteja no mesmo local de armazenamento do conteúdo audiovisual.

Um exemplo de aplicação envolvendo o MPEG-7 é o projeto *CESTA*. Esse projeto é comentado por Dallacosta, Souza e Tarouco (2004) e foi criado pelo *Centro Interdisciplinar de Novas Tecnologias na Educação* da UFRGS tendo como objetivo retornar vídeos educacionais, indexados em uma base de dados utilizando o MPEG-7, como resultados para um professor que utilizou uma palavra-chave em um sistema de busca.

Além disso, as descrições podem ser utilizadas independentemente de outros padrões MPEG tais como o MPEG-2 ou MPEG-4.

3.2.2 Elementos do MPEG-7

A criação das descrições utilizando o MPEG-7 é ressaltada na disponibilidade de elementos de metadados descritos em XML, *eXtreme Markup Languages*, suas estruturas e relacionamentos, além de possibilitar a especificação das descrições em formato binário.

As ferramentas de descrição formam o conjunto denominado *Description Tools* e utilizam as definições através de Descritores (*Descriptors*) e Esquemas de Descrição (*Description Schemes*) para alimentar as ferramentas de pesquisa, filtro e navegação nos conteúdos de multimídia (SILVA; GALINDO, 2007).

Descritores

Um descritor é um recurso de representação de uma característica específica de um objeto multimídia, geralmente apresentada sob a forma de um vetor numérico. Cada conteúdo numérico desse vetor representa uma observação extraída em um determinado intervalo de tempo. Descritores definem a sintaxe e a semântica da representação de recursos audiovisuais.

Usando descritores em uma abstração de baixo nível, eles podem representar a forma, movimento, textura, cor e movimento de câmera para objetos de vídeo ou imagens e também energia, harmonicidade e timbre para objetos de áudio, por exemplo. Enquanto descritores em um nível alto de abstração podem caracterizar eventos, conceitos abstratos, gêneros de conteúdo, etc.

Um exemplo de descritor para áudio é o *AudioPower*. Esse descritor descreve a potência instantânea de um sinal de áudio em um determinado instante de tempo.

A Figura 3.3 mostra o vetor de característica extraído do sinal de áudio pelo descritor *AudioPower* através de 03 segundos da música "*Breed*" da banda "*Nirvana*" utilizando os segundos iniciais a partir do primeiro minuto da música.

Para essa extração, foi configurado um intervalo de observação ou intervalo entre dois quadros consecutivos de 30 milissegundos (*hopSize*) gerando um vetor com 100 amostras

(*totalNumOfSamples*). A quantidade de amostras é obtida pela divisão do tempo total avaliado da música pelo *hopSize*.

```

▼<AudioDescriptor xsi:type="AudioPowerType">
  ▼<SeriesOfScalar hopSize="PT30N1000F" totalNumOfSamples="100">
    ▼<Raw>
      0.0011668458 0.024893103 0.06627376 0.026755724 0.02301172 0.018402187
      0.018366655 0.011658657 0.053358857 0.022628214 0.01470106 0.012289734
      0.015635913 0.011984479 0.04896287 0.026102716 0.027823081 0.021657227
      0.020333406 0.022824932 0.04033204 0.050043482 0.024386209 0.019392593
      0.017582582 0.017856786 0.016102249 0.06219893 0.026683645 0.017921235
      0.020135453 0.016461657 0.017924443 0.014871761 0.015792767 0.01771345
      0.014971545 0.014962951 0.014820441 0.031904604 0.025447987 0.014064249
      0.015426994 0.014211841 0.012476743 0.0065440843 0.009867979 0.011889882
      0.009798659 0.010244677 0.010439558 0.007874953 0.075180955 0.023216825
      0.016521899 0.011636874 0.011508957 0.016932793 0.06593054 0.015988471
      0.018606776 0.010771092 0.010917555 0.010736457 0.038518853 0.034218565
      0.016854664 0.017112993 0.016626006 0.0148798395 0.024956845 0.053330794
      0.012770215 0.014441604 0.012843073 0.01592417 0.018504309 0.057546955
      0.019930517 0.010735668 0.012207021 0.01594814 0.015551317 0.019420432
      0.017616147 0.015321669 0.013121741 0.016049411 0.013468867 0.016591664
      0.04079126 0.015555302 0.012553564 0.011770457 0.011615443 0.008324563
      0.009650531 0.013102424 0.0141256 0.012796812
    </Raw>
  </SeriesOfScalar>
</AudioDescriptor>

```

Figura 3.3: Vetor de Características do Descritor *AudioPower*

Os descritores MPEG-7 que pertencem ao nível baixo de abstração, *Low Level Descriptor*, podem representar a estrutura de suas descrições em dois tipos: *Series Of Scalar* ou *Series of Vectors*. Quando a descrição de uma característica vem representada por *Series Of Scalar*, como na Figura 3.3, significa que um único valor escalar representa a característica obtida no intervalo de tempo da observação (30ms).

Contudo, alguns descritores utilizam vetores numéricos para representar a característica obtida no intervalo de tempo da observação ao invés de um único valor escalar, utilizando em sua estrutura a definição de *Series of Vectors*. Nesse último caso, o descritor cria uma estrutura multi-dimensional ao colocar sub-vetores dentro do vetor que representa a descrição de sua característica.

A Figura 3.4 mostra a plotagem do vetor de características do descritor *AudioPower* para a música "*Breed*" do "*Nirvana*" e para a música "*Dia Azul*" de "*Djavan*" utilizando o software *Matlab*.

Através da análise da plotagem é possível identificar que a música do "*Nirvana*" apresenta picos de amplitude, eixo das ordenadas, mais elevados quando comparados aos picos da música do "*Djavan*" nos segundos avaliados.

Na Figura 3.4, o eixo das abscissas representa o número de amostras obtidas, onde no exemplo, para 03 segundos das músicas, foram obtidas 100 amostras. Dessa forma, é possível realizar um tipo de classificação musical baseada nesse descritor.

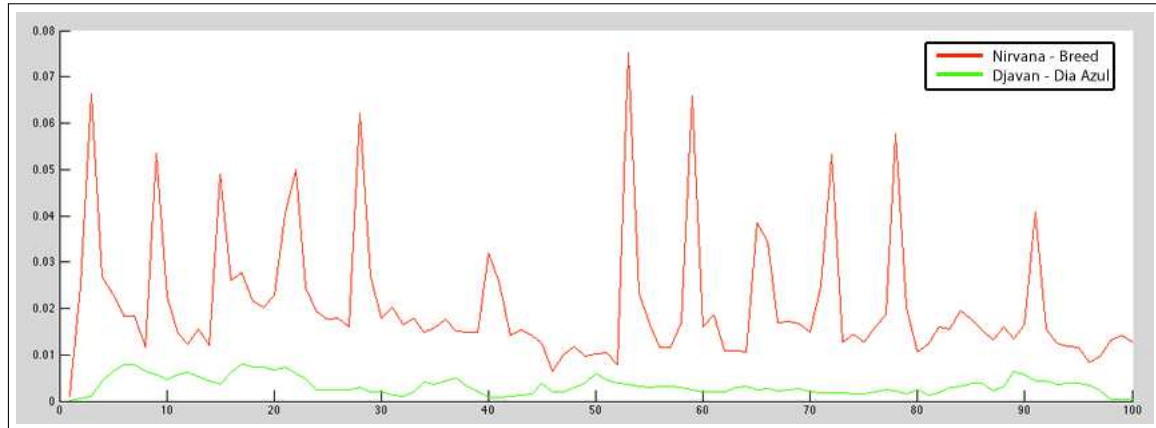


Figura 3.4: Plotagem do Descritor *AudioPower*.

Esquemas de Descrição

Os Esquemas de Descrição (*Description Schemes*), especificam a estrutura e a semântica das relações entre os seus componentes. Esses componentes podem ser Descritores ou outros Esquemas de Descrição. Uma Descrição consiste em um Esquema de Descrição e um conjunto de Descritores (instanciações) que descrevem um objeto multimídia.

A distinção entre um Esquema de Descrição e um Descritor é que um Descritor está preocupado com a representação de uma característica enquanto um Esquema de Descrição lida com a estrutura de uma Descrição. A Figura 3.5 mostra a estrutura de um Descritor e de um Esquema de Descrição e suas relações.

Linguagem de Definição de Descrição

No MPEG-7, a sintaxe de Descritores é definida pela Linguagem de Definição de Descrição (DDL), que é uma extensão da linguagem do esquema XML. A DDL permite a criação de novos Esquemas de Descrição e Descritores, e também permite a extensão e modificação de Esquemas de Descrição já existentes.

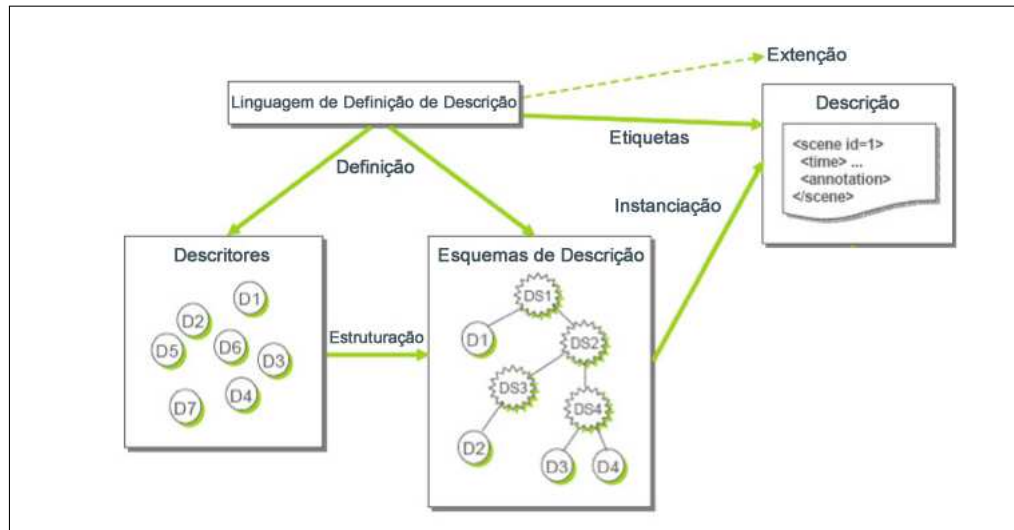


Figura 3.5: Modelo de Esquema de Descrição do MPEG-7.

Fonte: Adaptada de Dallacosta, Souza e Tarouco (2004).

A DDL é utilizada não só para definir a sintaxe dos descritores MPEG-7, mas também para permitir que os desenvolvedores declarem a sintaxe dos novos descritores que estão relacionados com as necessidades específicas da sua aplicação.

Assim, a DDL permite a criação de Ferramentas de Descrição, sejam Descritores ou Esquemas de Descrição, fornecendo recursos para as estruturas dos Descritores em Esquemas de Descrição.

A Figura 3.6 mostra um Esquema de Descrição e Descritores definidos por uma Linguagem de Definição de Descrição da música "Exodus" de "Bob Marley". Nessa Figura existem um conjunto de descritores, sendo que cada um é instanciado através do *AudioDescriptor*, e seus respectivos tipos de descrição.

3.2.3 Componentes do MPEG-7

Formalmente o MPEG-7 é constituído pela ISO 15938 possuindo os seguintes componentes: ISO/IEC 15938-1: MPEG-7 Systems, ISO/IEC 15938-2: *Description Definition Language*, ISO/IEC 15938-3: MPEG-7 Visual, ISO/IEC 15938-4: MPEG-7 Audio, ISO/IEC 15938-5: *Multimedia Description Schemes*, ISO/IEC 15938-6: *Reference Software* e ISO/IEC 15938-7: MPEG-7 Conformance.

```

▼<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  ▼<Description xsi:type="ContentEntityType">
    ▼<MultimediaContent xsi:type="AudioType">
      ▼<Audio xsi:type="AudioSegmentType">
        ▶<MediaInformation xsi:type="MediaInformationType">...</MediaInformation>
        ▶<AudioDescriptor xsi:type="AudioPowerType">...</AudioDescriptor>
        ▶<AudioDescriptor hiEdge="16000.0" loEdge="62.5" octaveResolution="1/0"
          xsi:type="AudioSpectrumBasisType">...</AudioDescriptor>
        ▶<AudioDescriptor xsi:type="AudioSpectrumProjectionType">...</AudioDescriptor>
        ▶<AudioDescriptor xsi:type="AudioSpectrumCentroidType">...</AudioDescriptor>
        ▶<AudioDescriptor hiEdge="16000.0" loEdge="62.5" octaveResolution="1/8"
          xsi:type="AudioSpectrumEnvelopeType">...</AudioDescriptor>
        ▶<AudioDescriptor hiEdge="16000.0" loEdge="250.0"
          xsi:type="AudioSpectrumFlatnessType">...</AudioDescriptor>
        ▶<AudioDescriptor xsi:type="AudioSpectrumSpreadType">...</AudioDescriptor>
        ▶<AudioDescriptor xsi:type="AudioWaveformType">...</AudioDescriptor>
        ▶<AudioDescriptionScheme xsi:type="AudioSignatureType">...
      </AudioDescriptionScheme>
    </Audio>
  </MultimediaContent>
</Description>
▼<Description xsi:type="ModelDescriptionType">
  ▶<Model numOfStates="12" xsi:type="SoundModelType">...</Model>
</Description>
</Mpeg7>

```

Figura 3.6: Exemplo de Esquema de Descrição do MPEG-7.

A ISO/IEC 15938-1: *MPEG-7 Systems* é uma padronização que especifica as funcionalidades a nível de sistema como a preparação de descrições MPEG-7 para transmissão e/ou armazenamento, sincronização de conteúdo e descrições, e desenvolvimento de decodificadores.

A ISO/IEC 15938-2: *Description Definition Language* é uma linguagem padronizada para definir novos DS's e D's ou para expandir ou modificar DS's ou D's já existentes.

A ISO/IEC 15938-3: *MPEG-7 Visual* especifica um conjunto de Descritores e Esquema de Descrição padronizados que dizem respeito somente a descrição visual. Descritores desse padrão abordam principalmente características específicas da imagem como cor, textura, forma e movimento.

A ISO/IEC 15938-4: *MPEG-7 Áudio* traz ferramentas de descrição para tratar somente de descrições de áudio sendo esse o componente principal aplicado a esse projeto. Utiliza estruturas que são um conjunto de descritores de baixo nível para características de áudio e também ferramentas de descrição de alto nível que são mais específicas para um conjunto de aplicações (ISO, 2002).

A ISO/IEC 15938-5: *Multimedia Description Schemes* (MDS) especifica uma estrutura de alto nível que permite a descrição genérica de todos os tipos multimídia, incluindo áudio,

visual e textual. Os MDS são estruturas de metadados para descrição de conteúdo de audiovisual.

A ISO/IEC 15938-6: *Reference Software* tem como objetivo fornecer uma referência de aplicação das partes relevantes do padrão MPEG-7, e é conhecido como software de experimentação. Tem status normativo.

A ISO/IEC 15938-7: *MPEG-7 Conformance* foi criado recentemente e tem como objetivo fornecer diretrizes e procedimentos para testar a conformidade das aplicações do MPEG-7.

Capítulo 4

Trabalhos Relacionados

Alguns trabalhos relacionados a recomendação de músicas são encontrados na literatura. Nesse Capítulo são apresentados trabalhos que representam sistemas de recomendação de músicas e de classificação de gêneros musicais.

O Apêndice B apresenta uma tabela resumo com as contribuições de cada trabalho relacionado e a adaptação do uso de recursos nesse projeto.

4.1 Classificação de Gêneros Musicais de Sinais de Áudio

Gêneros musicais são descrições categóricas usadas para caracterizar músicas em lojas, estações de rádio e na internet (TZANETAKIS; ESSL; COOK, 2002).

Para os autores desse projeto, Tzanetakis, Essl e Cook (2002), a divisão de músicas em gênero pode ser considerada subjetiva e arbitrária, porém existem critérios que podem ser utilizados durante a classificação para o processo ser mais preciso.

Alguns recursos como a percepção da textura, instrumentalização e estrutura rítmica podem ser utilizados para caracterizar um determinado gênero. Assim, nesse projeto é proposto um sistema de classificação automática de gênero musical utilizando algoritmos de classificação.

Esses algoritmos utilizam um conjunto de recursos que representam a superfície de uma

música e sua estrutura rítmica de sinais de áudio. Esse conjunto é avaliado através da formação de classificadores de reconhecimento de padrões estatísticos utilizando coleções de áudio.

Para isso, os autores propuseram utilizar um modelo gaussiano para representar cada classe como uma distribuição gaussiana simples com parâmetros estimados a partir do conjunto de dados de treino.

Assim, os sinais de áudio podem ser classificados automaticamente utilizando uma hierarquia de gêneros que podem ser representados como uma árvore de quinze nós. Para cada nó da árvore, foi treinado um conjunto de cinquenta amostras com trinta segundos cada.

A partir do nó inicial, o sinal de áudio de uma música é avaliado através de suas características se deslocando de um nó para outro até ser classificado em um determinado gênero musical. Proposta por Tzanetakis, Essl e Cook (2002), a Figura 4.1 apresenta um modelo da árvore de classificação automática de gêneros musicais.

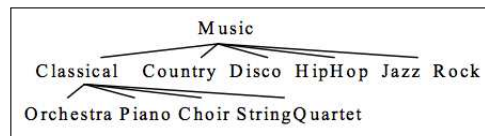


Figura 4.1: Modelo de Árvore para Classificação Automática de Gêneros Musicais
Fonte: Tzanetakis, Essl e Cook (2002).

4.2 Uma Pesquisa de Sistemas de Recomendação de Músicas e Perspectivas Futuras

Música é subjetiva e universal. A música não somente transmite emoção, mas pode indicar o humor do ouvinte (SONG; DIXON; PEARCE, 2012).

Para Song, Dixon e Pearce (2012), em um sistema de recomendação de conteúdo existem três principais componentes: usuários, itens e algoritmos para relacionar usuários e itens. Um sistema de recomendação eficiente necessita trabalhar com diversas preferências de um usuário, porém obter tais preferências é uma tarefa cara em termos de custos.

Diferenças entre regiões geográficas ou entre idades de usuários devem apresentar resultados de preferências musicais também diferentes. Assim, outros fatores como gênero, estilos de vida e interesses de um usuário também determinará suas escolhas por músicas. Ainda segundo Song, Dixon e Pearce (2012), diversas pesquisas tem revelado que inteligência, personalidade e preferências de um usuário estão ligadas.

Um perfil de um item a ser recomendado define diversas informações usadas em um processo de recuperação de informações de músicas. Podem ser analisados três perfis de metadados de uma música: metadados editoriais, metadados culturais e metadados acústicos.

Os metadados editoriais dizem respeito ao nome do compositor, título da música, gênero musical. Os metadados culturais são obtidos através da análise semântica da informação trazida na música como padrões adotados, categorias ou associações com fontes de documentos. E os metadados acústicos são obtidos pela análise do sinal de áudio como batidas, tempo, etc.

Song, Dixon e Pearce (2012) analisam diversos modelos de propostas para sistemas de recomendação de músicas e suas limitações. O modelo híbrido, por apresentar as vantagens de um sistema de recomendação baseado em conteúdo e também de um sistema com filtragem colaborativa, seria o melhor modelo proposto pelos autores, porém sua complexidade ainda não foi estudada completamente.

4.3 Nextone Player: Um Sistema de Recomendação de Músicas Baseado no Comportamento do Usuário

Hu e Ogihara (2011) propõem um sistema de recomendação de músicas que analisa o padrão de escuta de um usuário para estimar o seu nível de interesse para uma próxima música. Além disso, o comportamento do usuário para uma determinada música também é avaliado através de *feedback* para ajustar a estratégia de recomendação para a próxima música.

Esse *feedback* é obtido através da execução de uma música por um usuário. Se a música é escutada do início ao fim, o *feedback* é estabelecido positivamente, ou seja, o usuário gostou

da música. Porém, se a música é interrompida antes do minuto final de sua execução, o *feedback* é avaliado negativamente, ou seja, o usuário não gostou da música.

O sistema proposto pelos autores utiliza cinco perspectivas para saber se uma música está para ser recomendada: gênero, ano, contemporaneidade da música, a favorabilidade da música e o padrão de tempo. Cada perspectiva é calculada com pesos diferentes devido aos diferentes gostos musicais de diferentes usuários.

Utilizando o gênero e o ano como fatores de análise, foi notado que alguns usuários preferem ouvir músicas dentro de um mesmo gênero e ano musical, porém outros adoram ouvir músicas diversas de anos e gêneros diversificados. Assim, foi concluído por pelos autores desse projeto que, com uso desse dois fatores somente, não é possível realizar com eficiência uma boa opção de recomendação ao usuário.

Quantas vezes uma música foi ouvida e quantas vezes uma música foi ouvida completamente são dois fatores analisados para o estudo da favorabilidade de uma música. Ainda segundo Hu e Ogihara (2011), alguns usuários não gostam de ouvir a mesma música em pequenos intervalos de tempo, mesmo que ela seja uma de suas músicas favoritas.

Músicas que já foram favoritas para um usuário em um período passado, também podem ser recomendadas. A perspectiva da contemporaneidade analisa as músicas, que um dia foram favoritas, para recomendá-las no momento certo ao usuário.

A última perspectiva analisada é o padrão do tempo. Os usuários, por terem atividades diárias diferentes em um período diferente de um dia ou uma semana, tendem a selecionar estilos de músicas diversificados. Por exemplo, no período da tarde, um usuário pode gostar de um tipo de música calma, mas pode mudar para músicas mais agitadas durante a noite.

Um modelo de mistura gaussiana é utilizado por esse projeto para representar o padrão de tempo de escuta e também calcular a probabilidade de tocar uma música em um determinado momento.

4.4 Classificação de Áudio Baseada em Extração de Características e Redes Neurais

O trabalho proposto por Fachini (2011) apresenta um estudo realizado sobre técnicas de extração de características de áudio que descrevem o timbre dos sons e também o uso de Redes Neurais Artificiais com objetivo de desenvolver um sistema que facilite a classificação, indexação e recuperação automática de *samples* de instrumentos musicais. Em música, *sample* é um trecho de áudio gravado utilizado para composição musical, geralmente com o auxílio de um programa ou equipamento chamado sampler.

Esse trabalho realizou extração de características de áudio armazenando-as em um vetor de características que são formados por valores numéricos. O autor, para extrair essas características, utilizou análise de tempo e frequência através de técnicas como a Transformada de Fourier, aplicações de filtros e escalas.

A variação da amplitude das diferentes frequências geradas pelos instrumentos musicais é o que dará ao instrumento sua característica sonora, o timbre. Para descrever o timbre, o autor desse projeto utilizou dois vetores de características. Um vetor de característica foi extraído utilizando os coeficientes *Mel-Frequency Cepstral* e o outro vetor foi gerado utilizando os descritores de timbre do padrão MPEG-7.

Os Coeficientes Mel-Cepstrais (*Mel-Frequency Cepstral Coefficients - MFCC*) são tipicamente utilizados para tarefas de reconhecimento de fala e que também vêm sendo utilizados em aplicações musicais. Esse uso crescente desses coeficientes ocorre devido ao fato de terem sido desenvolvidos para modelar o envelope espectral do sinal de uma forma muito compacta levando em conta a percepção não linear do som pelo ouvido humano.

Assim, todo o processo envolvido na geração do vetor de características utilizando esses coeficientes baseia-se em estudos na área de psicoacústica, que estuda a percepção auditiva humana.

Na área do MPEG-7 foram utilizados os descritores de timbre que visam descrever as características perceptuais do som. O timbre é tratado como um vetor de características

que inclui o envelope temporal e espectral e variações destes.

Os descritores de timbre temporal são extraídos do envelope de amplitude de um sinal no domínio do tempo. O envelope de amplitude do sinal descreve a variação de energia do sinal em vários estágios e pode ser representado pelo chamado *ADSR* (*Attack, Decay, Sustain, Release*).

A Figura 4.2 apresenta um modelo de aproximação linear do envelope de um sinal no qual é possível observar os estágios do *ADSR*.

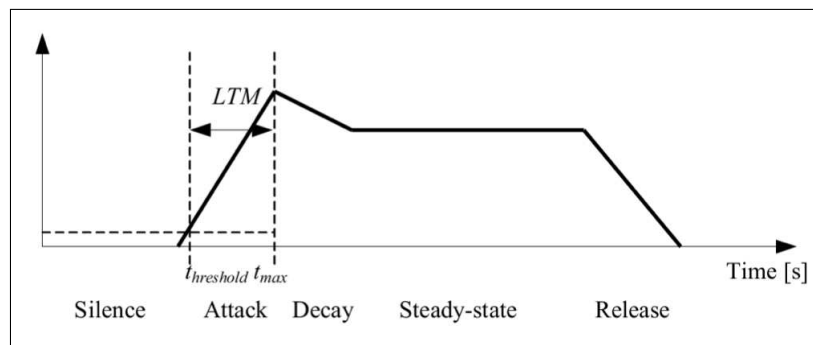


Figura 4.2: Aproximação Linear do Envelope de Um Sinal de Áudio
Fonte: Fachini (2011).

Fachini (2011) utilizou dois descritores de timbre temporal: *Log Attack Time* e o *Temporal Centroid*. O *Log Attack Time* é o descritor de ataque que indica o tempo que se leva para a progressão de um nível de energia limiar para o nível de energia máximo do envelope de amplitude do sinal.

Já o *Temporal Centroid* caracteriza o envelope do sinal e é definido como o tempo médio sobre o envelope de amplitude do sinal, representando em que ponto a energia do sinal se concentra e possibilita distinguir entre sons com ataque similar, mas com tempos de decaimento e suspensão diferentes.

Para a extração dos descritores de timbre espectrais é necessário estimar a frequência fundamental f_0 e detectar os componentes harmônicos do sinal. O padrão MPEG-7 não especifica como extrair estas informações do sinal.

As redes neurais foram propostas por Fachini (2011) como solução para um sistema de classificação automática de áudio baseado em extração de características.

O sistema foi dividido em quatro partes principais: criação de um banco de sons de instrumentos musicais, extração de características, sistemas de classificação e sons classificados.

Os vetores de características foram gerados a partir de áudio produzido por instrumentos musicais percussivos e a linguagem Python foi usada para o desenvolvimento do sistema. O autor utilizou, ao todo, 279 *samples* para os testes.

Após os testes, foi concluído por Fachini (2011) que enquanto a classificação utilizando MFCC possui como entrada um vetor de 40 coeficientes, apenas 7 são utilizados para a classificação utilizando os descritores de timbre do padrão MPEG-7.

Dessa forma, pode-se concluir que o MFCC descreve com maior precisão sons de instrumentos musicais. O seu vetor de características mostrou-se mais eficiente para descrever o timbre de um instrumento musical em comparação aos descritores de timbre temporal e espectral propostos pelo padrão MPEG-7.

Capítulo 5

Arquitetura do Sistema Para Uso Individual de Metadados

O uso de reprodutores de arquivos multimídia tem aumentado consideravelmente nos últimos anos. O gerenciamento e recuperação de arquivos de áudio em uma base de dados tem sido o grande desafio frente a esse crescimento.

Fornecer ao usuário um acesso rápido aos arquivos audiovisuais de seu interesse, em meio a uma grande quantidade de arquivos disponibilizados, tem sido o objetivo de sistemas de recomendação. Portanto, esse capítulo propõe uma arquitetura para um sistema de recomendação de músicas baseado em gêneros musicais que utilize, separadamente, as etiquetas ID3 e os descritores do MPEG-7 para gerar listas de recomendação.

Esse Capítulo é dividido em três seções. Na seção 5.1 é apresentada a arquitetura desse sistema. Na seção 5.2 são caracterizados os métodos de extrações das etiquetas ID3 e dos descritores do MPEG-7 e mais as implementações dos módulos de banco de dados e de recomendação. E na seção 5.3 são apresentados os resultados e as análises experimentais.

5.1 Arquitetura

A Figura 5.1 representa a arquitetura que foi desenvolvida para o sistema de recomendação de músicas baseado em gêneros musicais que utiliza de forma individual as etiquetas ID3 e os descritores do MPEG-7. Essa arquitetura é composta de dois módulos, um de banco de dados e o outro de recomendação.

Não é possível a utilização dessa arquitetura com a finalidade de melhorar a eficiência no processo de recomendação de músicas utilizando mais de um descritor do MPEG-7 ou integrando as etiquetas ID3 com os descritores do MPEG-7. Para essa finalidade são necessárias alterações na implementação as quais são apresentadas no próximo capítulo.

Nesse sistema, o banco de dados contém uma tabela com o nome de todas as músicas disponíveis. Através dessa tabela, uma lista de reprodução é gerada automaticamente e apresentada ao usuário com todas as músicas disponíveis para serem ouvidas. Essa etapa está representada pelo fluxo 1.

O fluxo 2 ocorre quando a música é selecionada pelo usuário e então o módulo de banco de dados recebe o ID dessa música. O fluxo 3 e o fluxo 4 representam consultas no banco de dados utilizando, como parâmetros, os metadados da música selecionada através do ID e o retorno dessas informações, respectivamente.

O fluxo 5 representa a formação das listas de músicas através dos retornos das consultas no banco de dados. Para a formação da lista de músicas utilizando os descritores MPEG-7 há uma necessidade de repetir os processos representados pelos fluxos 3 e 4 no módulo de banco de dados, de acordo com a quantidade de músicas presentes na base de dados devido às leituras nos arquivos XML de cada música. Essas leituras serão vistas posteriormente nos algoritmos que compõe cada módulo. Para cada descritor do MPEG-7, o processo de recomendação

Por fim, o fluxo 6 representa a entrada das duas relações de músicas candidatas no módulo de recomendação sendo que uma relação foi gerada utilizando o ID3 e a outra um descritor do MPEG-7. O módulo de recomendação realiza toda a atividade pertinente à escolha das músicas a serem recomendadas que, através do fluxo 7, são apresentadas ao usuário.

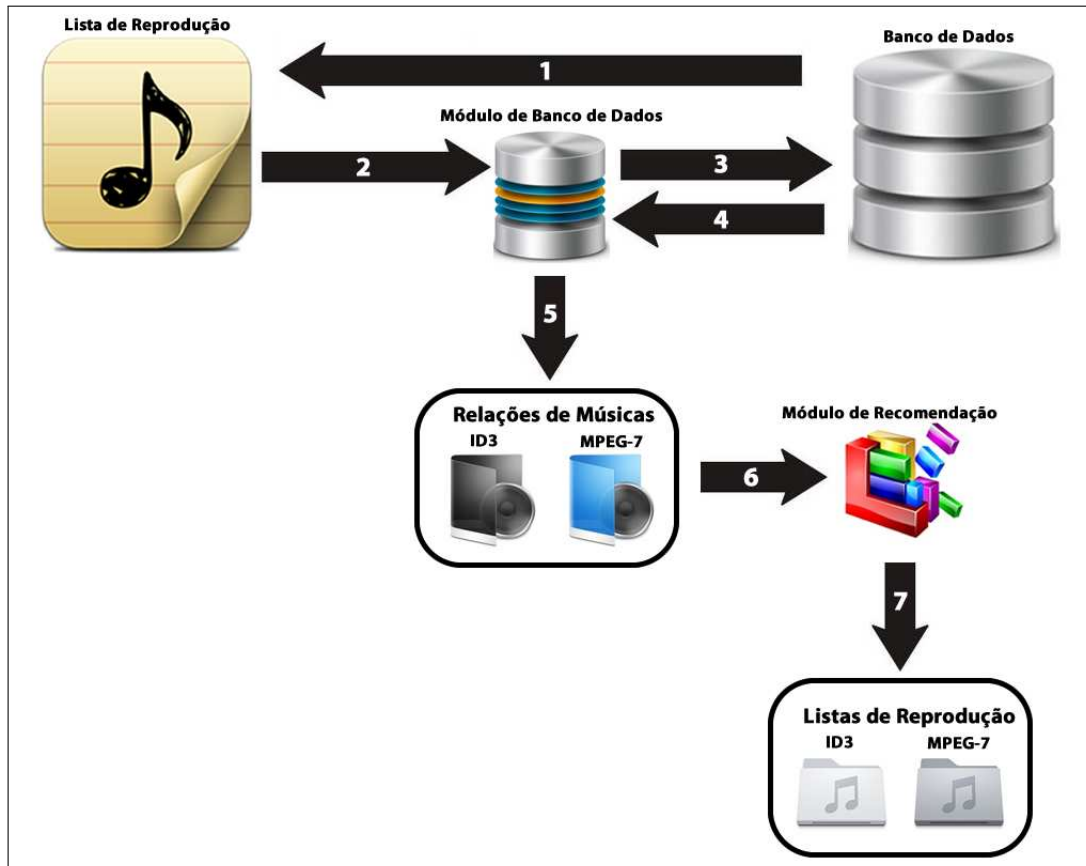


Figura 5.1: Arquitetura do Sistema.

5.2 Implementação

A implementação desse sistema ocorre em quatro etapas. A primeira etapa é formada pelo método de extração das etiquetas ID3 presentes nos arquivos de músicas utilizados nesse projeto. A segunda etapa é formada pelo método de extração dos descritores do MPEG-7. E as duas últimas etapas são formadas pela implementação do módulo de banco de dados e do módulo de recomendação, respectivamente.

5.2.1 Método de Extração das Etiquetas ID3

Os arquivos MP3, além da música propriamente dita, trazem em sua estrutura metadados que classificam a música na qual estão inseridos.

Assim, foi necessário o desenvolvimento de um algoritmo que fizesse a leitura dos metadados

contidos nos arquivos MP3 extraíndo-os para uso posterior nos módulos do sistema. A linguagem *Java* foi escolhida para realizar essa leitura.

Para a construção da base de dados, vista com mais detalhes posteriormente, foram utilizados mil arquivos MP3. Dessa forma, havia a necessidade desse algoritmo ler todos os mil arquivos e extrair as etiquetas necessárias para a construção da tabela do banco de dados.

Esse projeto utilizou somente a etiqueta que traz informações sobre o gênero da música, visto que a recomendação de músicas desse projeto ocorre apresentando ao usuário músicas do mesmo gênero musical de sua preferência com base na música escolhida por ele na reprodução.

Foi utilizada uma biblioteca em Java chamada *mp3agic* responsável por ler arquivos mp3 e manipular suas etiquetas.

Para esse projeto, somente a leitura foi utilizada como operação, excluindo-se qualquer outra que pudesse modificar o conteúdo existente nas informações das etiquetas.

Nesse contexto, o Algoritmo 5.1 apresenta a classe *Musica* e todos os atributos que o objeto *musica* irá ter. Foi incluído o atributo *nome*, visto nas linhas 6 à 10 e o atributo *local*, visto nas linhas 18 à 22. Esse último atributo é responsável por conter o local onde a música está armazenada.

O Algoritmo 5.2 representa a classe responsável pela leitura das etiquetas dos arquivos MP3 utilizando a biblioteca citada acima. Esse algoritmo apresenta em sua linha 8 a criação do objeto *musica*.

Esse objeto vai receber todos os atributos que posteriormente serão gravados no banco de dados a partir do objeto *dao*, linha 9, e que cuja classe será vista posteriormente. Para esse projeto, os mil arquivos MP3 foram colocados em uma única pasta.

Na linha 10, o endereço dessa pasta é obtido para que seja efetuada a leitura dos arquivos. A partir da leitura de cada arquivo, as informações referentes ao nome da música e sua localização são inseridas no objeto, conforme as linhas 12 à 18.

A partir da linha 19 ocorre a leitura da etiqueta ID3 que faz referência a informação de

Algoritmo 5.1 Algoritmo da Classe Musica

```
1: public class Musica {
2:     public String nome;
3:     public String genero;
4:     public String local;
5:
6:     public String getNome() {
7:         return nome;
8:     }
9:     public void setNome(String nome) {
10:        this.nome = nome;
11:    }
12:    public String getGenero() {
13:        return genero;
14:    }
15:    public void setGenero(String genero) {
16:        this.genero = genero;
17:    }
18:    public String getLocal() {
19:        return local;
20:    }
21:    public void setLocal(String local) {
22:        this.local = local;
23:    }
24: }
```

gênero. Essa informação, algumas vezes, vem suprimida no arquivo, sendo que para esses casos foi colocado um hífen no registro da variável do objeto, linhas 20 à 24.

Já na linha 25, há a passagem do objeto *musica* para a classe *MusicaDAO* para ser possível a leitura de seus atributos e o devido registro no banco de dados.

Dessa forma, após todos os arquivos terem seus atributos lidos e seus objetos criados, as informações são inseridas no banco de dados, finalizando a extração da Etiqueta ID3.

5.2.2 Método de Extração dos Descritores MPEG-7

O padrão MPEG-7 não define um método específico de como realizar a descrição de um arquivo multimídia dentro do contexto de sua norma, deixando o usuário com liberdade para usar quaisquer ferramentas que atendam a padronização.

Assim, foi utilizado nesse projeto um framework chamado *MPEG-7 Audio Encoder* para realizar descrições das músicas através dos descritores do padrão MPEG-7. Com o uso

Algoritmo 5.2 Algoritmo da Extração das Etiquetas ID3

```
1: import java.io.IOException;
2: import com.mpatric.*;
3: import com.mpatric.mp3agic.*;
4: import java.io.File;
5: public class Id3 {
6:     public static void main (String[] args) throws
7:     UnsupportedOperationException, InvalidDataException, IOException{
8:         Musica musica = new Musica();
9:         MusicaDAO dao = new MusicaDAO();
10:        File folder = new File("/Applications/MAMP/htdocs/tcc/musicas");
11:        File[] listOfFiles = folder.listFiles();
12:        for (int i = 1; i < listOfFiles.length; i++) {
13:            if (listOfFiles[i].isFile()) {
14:                Mp3File mp3file = new Mp3File("/Applications/MAMP/htdocs/tcc/
15:                musicas/"+listOfFiles[i].getName());
16:                musica.setNome(listOfFiles[i].getName());
17:                musica.setLocal("localhost/tcc/musicas/"+listOfFiles[i].getName()
18:                +".mp3");
19:                ID3v2 id3v2Tag = mp3file.getId3v2Tag();
20:                if(id3v2Tag.getGenreDescription()==null){
21:                    musica.setGenero("-");
22:                }else{
23:                    musica.setGenero(id3v2Tag.getGenreDescription());
24:                }
25:                dao.adiciona(musica);
26:            }
27:        }
28:    }
29: }
```

desses descritores, esse framework gera o esquema de descrição para uma determinada música salvando-o em formato XML.

No final do processo de extração das descrições de todas as músicas desse projeto, haverá mil arquivos XML sendo que cada um deles pertencerá exclusivamente a uma determinada música.

Para realizar a extração das características através do MPEG-7, as músicas, por serem advindas de diversas fontes, sofreram dois ajustes de padronização. Assim como Hellmuth et al. (2001) e Allamanche et al. (2001) propuseram, foram ajustados os canais de áudio das músicas para mono com a finalidade de se obter uma padronização no sinal.

O segundo ajuste foi padronizar o tempo usado para extrair as características das músicas. Esse ajuste foi necessário devido ao fato das músicas terem tempos de duração diferentes.

Assim, para uma música de 05 minutos haveria um vetor de características muito maior quando comparado a um vetor de características extraído de uma música de 03 minutos usando um mesmo descritor.

Huang e Jenor (2004) e Pampalk (2006) utilizaram 30 segundos iniciais de cada áudio para extrair as características, porém foi observado que os segundos iniciais das músicas podem não conter a música propriamente dita ou até mesmo conter, nos segundos iniciais, intervalos sem áudio.

Além disso, Pampalk (2006) ainda ressalta que um tempo menor que 30 segundos dificulta o trabalho de análise de similaridade devido às poucas informações extraídas.

Assim, para esse projeto, foi estabelecida a padronização de tempo de 60 segundos, contados a partir do primeiro minuto, para serem utilizados pelos descritores do MPEG-7 na extração de características.

Os Descritores do Padrão MPEG-7 Utilizados

Foram utilizados oito descritores de baixo nível para formar um esquema de descrição e para alimentar os métodos de comparação entre as músicas: *Audiopower*, *Waveform*, *Audio Spectrum Flatness*, *Audio Spectrum Centroid*, *Audio Spectrum Spread*, *Audio Spectrum Envelope*, *Audio Spectrum Basis* e o *Audio Spectrum Projection*.

A escolha de descritores de baixo nível foi feita pelo fato desses descritores, de forma ímpar, definirem conceitos sobre uma determinada característica sem a necessidade de envolver um conjunto completo de eventos.

O *Audiopower* permite medir a evolução da amplitude do sinal em função do tempo. Quanto maior é a amplitude, mais forte é o som. Assim, o *AudioPower* descreve a potência instantânea suavizada do sinal de áudio. Em um gráfico desse descritor, os picos do sinal correspondem às partes onde o sinal original tem a maior amplitude.

O *Waveform* é uma descrição compacta de um sinal de áudio considerando os pares formados por valores mínimos e máximos em um determinado tempo. Esse descritor fornece uma estimativa do sinal de áudio no domínio do tempo, além de permitir o armazenamento

econômico e simples, exibir ou comparar as técnicas de formas de onda.

Esses dois descritores são básicos do MPEG-7, pois promovem uma simples e econômica descrição das propriedades temporais de um sinal de áudio. A Figura 5.2 mostra a representação de um sinal original de uma música e as representações do *Audiopower* e do *Waveform*.

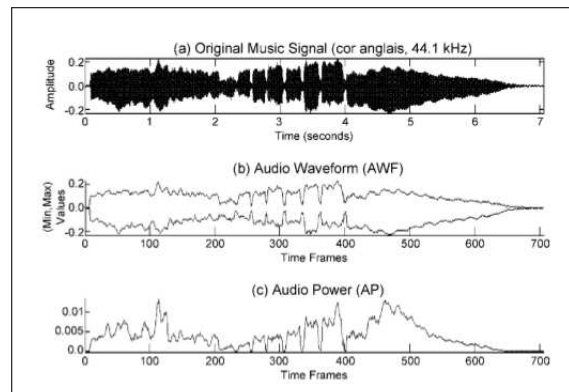


Figura 5.2: Representação dos Descritores *Audiopower* e *Waveform*.

Fonte: Kim, Moreau e Sikora (2006).

O *Audio Spectrum Flatness* descreve as propriedades de nivelamento do espectro de energia de um sinal de áudio dentro de um determinado número de faixas de frequências. Cada valor obtido expressa o desvio do espectro de potência do sinal a partir de um estado plano.

O número de faixas de frequências varia de um *loEdge* para um *hiEdge* que representam a menor e a maior frequências da faixa. Kim, Moreau e Sikora (2006) propõem que a frequência mais baixa seja de 250Hz e a mais alta não ultrapasse a frequência de Nyquist (o dobro da maior frequência do sinal).

O *Audio Spectrum Centroid* é o centro da gravidade de uma frequência logarítmica de um espectro de potência. Ele contém informações perceptuais de brilho que indica se o espectro de potência é dominado por altas ou baixas frequências. Esse descritor aproxima as frequências de percepção para o sistema de audição humana, tendo ligação com a nitidez de uma música.

O *Audio Spectrum Spread* usa a largura de banda de frequências de um determinado tempo para apresentar indicações de como o espectro é distribuído ao redor da *centroid*. Baixos valores significam que o espectro se concentra ao redor da *centroid* e altos valores refletem a

distribuição do poder através de uma ampla gama de frequências. Esse descritor permite a diferenciação entre sons de tons semelhantes e de ruídos semelhantes.

O *Audio Spectrum Envelope* é o espectro de potência de frequência logarítmica que pode ser usado para gerar um reduzido espectrograma do sinal de áudio original. Esse descritor pode ser usado para exibir um espectrograma, para sintetizar uma auralização dos dados ou como um descritor de utilização geral para a busca e comparação (ISO, 2002). A Figura 5.3 mostra a representação dos descritores espectrais usados nesse projeto.

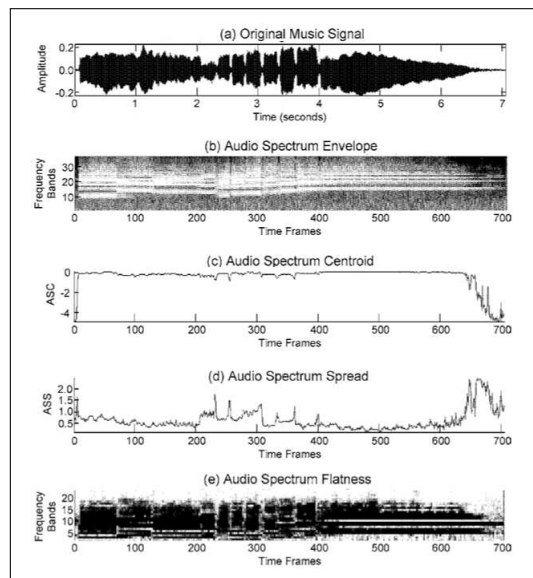


Figura 5.3: Representação dos Descritores Espectrais.

Fonte: Adaptada de Kim, Moreau e Sikora (2006).

O *Audio Spectrum Basis* contém funções de base que são usadas para projetar as descrições do espectro de alta dimensão para uma representação de baixa dimensão adequada para um modelo de classificadores. Dessa forma, o teor máximo de informações do espectro é mantido mesmo reduzindo a dimensionalidade.

A redução de dimensionalidade desempenha um papel substancial em aplicações de classificação automática por representar de forma compacta a informação saliente sobre os segmentos de áudio. Estas características têm demonstrado um bom desempenho em aplicações de classificação e recuperação automática.

O *Audio Spectrum Projection* é o complemento ao *Spectrum Basis* e é utilizado para representar características de baixa dimensionalidade de um espectro após a projeção sobre

uma base de classificação reduzida.

Esses dois últimos descritores são utilizados na classificação de som e ferramentas de indexação para a classificação automática de segmentos de áudio, utilizando modelos probabilísticos. Um exemplo de aplicação é a classificação automática de gêneros musicais (ISO, 2002).

5.2.3 Módulo de Banco de Dados do Sistema

Inicialmente, o banco de dados foi composto por três tabelas sendo que a coluna de ID, existente nas três, é igual em todas elas. Além disso, cada ID referencia sempre a mesma música em qualquer uma das tabelas.

A tabela *Playlist*, composta pelas colunas *id*, *nome* e *LocalMusica*, é a tabela que gera a lista de reprodução inicial através de comandos do MySQL. Assim, o usuário obtém uma lista de reprodução com todos os títulos contidos na base de dados disponíveis para serem reproduzidos.

Através do ID da música escolhida pelo usuário nessa lista de reprodução, é possível obter o seu respectivo gênero através de uma consulta na tabela *Etiquetas* e filtrar todas as outras músicas que contenham o mesmo gênero musical, obtendo dessa forma, a primeira relação de músicas utilizando as etiquetas ID3.

A tabela *Etiquetas* é composta pelas colunas *id*, *nome* e *genero* e é alimentada pelo algoritmo Java explicado na Subseção 3.2.1.

Essa primeira relação de músicas é processada pelo módulo de recomendação para gerar a lista de reprodução com as músicas que serão recomendadas ao usuário.

O Algoritmo 5.3 apresenta o código na linguagem PHP responsável pela obtenção do ID da música escolhida e pela filtragem no banco de dados utilizando o gênero da música selecionada.

Realizando uma análise do Algoritmo 5.3, é encontrado na linha 1 o método de recebimento do ID da música selecionada. Através do ID, há um procedimento de obtenção do gênero da

Algoritmo 5.3 Algoritmo do Módulo de Banco de Dados do ID3

```
1:  $id_escolha=$_GET['id'];
2:
3:  $sql = "SELECT * FROM etiquetas where id=$id_escolha";
4:  $query = mysql_query($sql);
5:  $resultado = mysql_fetch_assoc($query);
6:  $genero = $resultado['genero'];
7:  $nome = $resultado['nome'];
8:  $id = $resultado['id'];
9:
10: $query = mysql_query("SELECT * FROM etiquetas where genero='$genero'");
11: $num = mysql_num_rows($query);
12:
```

música escolhida, visto nas linhas 3 a 8. Nas linhas 10 a 12 ocorre a filtragem no banco de dados utilizando esse determinado gênero finalizando o procedimento do módulo de banco de dados para as Etiquetas ID3.

A terceira tabela criada é a tabela *mpeg7*. Essa tabela é composta pelas colunas *id*, *nome* e *localXML* sendo que essa última coluna contém o local de armazenamento do arquivo XML que, por sua vez, contém toda a descrição da música representada pelo padrão MPEG-7.

Com o ID da música selecionada pelo usuário, o sistema abre o arquivo XML correspondente. Após a abertura desse arquivo, o sistema precisa realizar a leitura dos arquivos XML das demais músicas que compõem a base de dados e medir o grau de similaridade dessas músicas com a música selecionada inicialmente utilizando um determinado descritor.

Assim como Eidenberger (2003) defende em seu artigo a distância euclidiana como um recurso de cálculo de similaridade dentro do padrão MPEG-7, esse método foi utilizado nesse projeto. Sendo $P = (p_x, p_y)$ e $Q = (q_x, q_y)$, a distância euclidiana pode ser calculada entre esses pontos utilizando a fórmula representada na Figura 5.4.

$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2} = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}.$$

Figura 5.4: Fórmula do Cálculo da Distância Euclidiana.

Se faz necessário antes de calcular a distância euclidiana realizar a normalização de cada

vetor de características que compõe cada descritor para que seja encontrado o seu respectivo vetor unitário.

A normalização é necessária para que os vetores correspondam a pontos dentro do espaço euclidiano para que seja calculada a distância entre eles.

Após o processo de normalização, o vetor de características que compõem cada descritor é utilizado de forma individual para calcular a distância euclidiana entre a música selecionada e as demais músicas da base de dados. Cada descritor do MPEG-7 é processado de forma individual e em sequência.

Dessa forma, ao final do cálculo de todas as distâncias euclidianas dos oito descritores, existirão oito relações independentes de músicas candidatas contendo o grau de similaridade entre a música selecionada e as demais músicas da base de dados.

Esse grau de similaridade é obtido pelos resultados das distâncias euclidianas calculadas. Assim, existirá uma relação de músicas candidatas para o descritor *AudioPower*, por exemplo, outra para o descritor *Waveform* e outras para os outros descritores.

Essas relações contém as músicas em ordem crescente pelo resultado do cálculo da distância euclidiana. A primeira música de cada relação possui uma distância euclidiana menor para a música selecionada pelo usuário, quando comparada às demais músicas, por exemplo. Esse ordenamento de cada relação de músicas foi obtido através do uso de funções e vetores em PHP.

O Algoritmo 5.4 representa o código da linguagem PHP que processa os cálculos da distância euclidiana para um determinado descritor, *Spectrum Spread*, e a geração de um vetor com os resultados desses cálculos.

Através da análise desse algoritmo, responsável pela geração do vetor com os valores das distâncias euclidianas, é visto que nas linhas 3 a 9 há um procedimento responsável pela obtenção do endereço do local onde o XML está armazenado e na linha 10 ocorre a leitura desse arquivo.

Após obter e ler o XML, o algoritmo, nas linhas 12 a 14, processa o vetor de características do descritor utilizado da música selecionada pelo usuário.

Algoritmo 5.4 Algoritmo do Módulo de Banco de Dados do MPEG-7

```

1:$vetor = Descritor_SpectrumSpread($id_escolha);
2:
3:function Descritor_SpectrumSpread($id_escolha){
4: $sql = "SELECT * FROM mpeg7 where id=$id_escolha";
5: $query = mysql_query($sql);
6: $resultado = mysql_fetch_assoc($query);
7: $xml_url = $resultado['localXML'];
8: $http = "http://";
9: $url_1 = $http.$xml_url;
10: $xml_1 = simplexml_load_file($url_1);
11:
12: $SpectrumSpread_1 = $xml_1->Description[0]->MultimediaContent->Audio->
13: AudioDescriptor[6]->SeriesOfScalar->Raw;
14: $SpectrumSpread_array_1 = explode(" ", $SpectrumSpread_1);
15: for($id=1;$id<1001;$id++){
16:   $SpectrumSpread_array_1_aux = $SpectrumSpread_array_1;
17:   $sql = "SELECT * FROM mpeg7 where id=$id";
18:   $query = mysql_query($sql);
19:   $resultado = mysql_fetch_assoc($query);
20:   $xml_url = $resultado['localXML'];
21:   $nome = $resultado['nome'];
22:   $http = "http://";
23:   $url_2 = $http.$xml_url;
24:   $xml_2 = simplexml_load_file($url_2);
25:   $SpectrumSpread_2 = $xml_2->Description[0]->MultimediaContent->Audio->
26:   AudioDescriptor[6]->SeriesOfScalar->Raw;
27:   $SpectrumSpread_array_2 = explode(" ", $SpectrumSpread_2);
28:   $SpectrumSpread_array_2_aux = $SpectrumSpread_array_2;
29:   $cont = count($SpectrumSpread_array_1);
30:   for($i =0; $i < $cont; $i++){
31:     $SpectrumSpread_array_1_aux[$i]= $SpectrumSpread_array_1_aux[$i]*
32:     $SpectrumSpread_array_1_aux[$i];
33:     $SpectrumSpread_array_2_aux[$i]= $SpectrumSpread_array_2_aux[$i]*
34:     $SpectrumSpread_array_2_aux[$i];
35:   }
36:   $SpectrumSpread_array_1_Norm= array_sum($SpectrumSpread_array_1_aux);
37:   $SpectrumSpread_array_2_Norm= array_sum($SpectrumSpread_array_2_aux);
38:   $SpectrumSpread_array_1_Norm = sqrt($SpectrumSpread_array_1_Norm);
39:   $SpectrumSpread_array_2_Norm = sqrt($SpectrumSpread_array_2_Norm);
40:   for($i =0; $i < $cont; $i++){
41:     $SpectrumSpread_array_1_Normalized[$i] = $SpectrumSpread_array_1[$i]/
42:     $SpectrumSpread_array_1_Norm;
43:     $SpectrumSpread_array_2_Normalized[$i] = $SpectrumSpread_array_2[$i]/
44:     $SpectrumSpread_array_2_Norm;
45:     $dist_euclid[$i] = $SpectrumSpread_array_1_Normalized[$i] -
46:     $SpectrumSpread_array_2_Normalized[$i];
47:   }
48:   for($i =0; $i < $cont; $i++){
49:     $dist_euclid[$i] = $dist_euclid[$i]*$dist_euclid[$i];
50:   }
51:   $dist_euclid_result= array_sum($dist_euclid);
52:   $dist_euclid_result = sqrt($dist_euclid_result);
53:   $result_aux[$indice] = $dist_euclid_result;
54:   $indice++;
55: }
56: return $result_aux;
57:}

```

Da linha 15 a 55 há um laço de repetição que possibilita a leitura de todos os XML's das demais músicas e os cálculos necessários para o processamento da distância euclidiana. Dentro desse laço, na linha 29, ocorre a verificação de quantos valores existem no vetor de características do descritor utilizado para que posteriormente ocorra o início da normalização vetorial, pois a quantidade desses valores variam a cada descritor do MPEG-7.

Nas linhas 30 a 47 ocorre a normalização vetorial, tendo início com a obtenção do quadrado de todos os valores dos vetores de características. Posteriormente, com a soma desses valores e o cálculo da raiz quadrada dessa soma, se obtém a norma vetorial. Com a norma de cada vetor, a normalização termina pela divisão de cada valor que compõem cada índice do vetor pela sua respectiva norma, linhas 40 a 44.

Então, a distância euclidiana é calculada e o valor é armazenado em um vetor, linhas 45 a 53. Por fim, essa função é finalizada com o retorno desse vetor contendo todos os resultados das distâncias euclidianas, linha 56.

Cada posição desse vetor tem relação direta com o ID das músicas da base de dados. A primeira posição desse vetor, por exemplo, recebe o resultado da distância euclidiana calculado entre música de ID número 1 da base de dados e a música selecionada pelo usuário. A segunda posição recebe o resultado envolvendo a música de ID 2 e a música selecionada. E assim sucessivamente.

5.2.4 Módulo de Recomendação

Esse módulo tem como objetivo processar as listas geradas pelo módulo do banco de dados e realizar a recomendação das músicas, sendo que serão recomendadas dez músicas para cada um dos metadados.

A lista gerada pelo uso das Etiquetas ID3 pode conter mais que dez títulos musicais com o mesmo gênero da música selecionada pelo usuário, tendo cada uma dessas músicas pesos iguais de recomendação.

Por isso, foi utilizada uma estrutura no código PHP que, de forma aleatória e sem repetição, tenha a função de gerar uma lista de reprodução com no máximo dez títulos.

Algoritmo 5.5 Algoritmo do Módulo de Recomendação do ID3

```

1: $array_ids = array();
2: $array_nomes = array();
3: $array_generos = array();
4: $cont=0;
5:
6: echo "<table width='70%' border='0' align='center'>
7: <tr>
8: <td width='8%' align='left' valign='middle'><font color='red'><b>ID
9: </b></font></td>
10: <td width='90%' align='left' valign='middle'><font color='red'><b>
11: NOME DA MUSICA</b></font></td>
12: <td width='90%' align='left' valign='middle'><font color='red'><b>
13: GÊNERO</b></font></td>
14: </tr>";
15:
16: while ($prod = mysql_fetch_array($query)) {
17:   $array_ids[$cont]=$prod[0];
18:   $array_nomes[$cont]=$prod[1];
19:   $array_generos[$cont]=$prod[2];
20:   $cont++;
21: }
22: $cont=0;
23: $array_verifica = array();
24:
25: while (($cont<$num) && ($cont<10)) {
26:   $ran = rand(0,$num-1);
27:   if (!in_array($array_ids[$ran], $resultID3)) {
28:     $resultID3[$cont]=$array_ids[$ran];
29:     echo "<tr>
30: <td align='left' valign='middle'>".$array_ids[$ran]."</td>
31: <td align='left' valign='middle'><a href='musica.php?id=
32: $array_ids[$ran]'>".$array_nomes[$ran]."</a></td>
33: <td align='left' valign='middle'>".$array_generos[$ran]."</td>
34: </tr>";
35:
36:   $cont++;
37: }
38: }
39: echo "</table>";

```

O Algoritmo 5.5 apresenta o código de processamento da lista gerada pelas etiquetas ID3. Nesse algoritmo ocorre a criação da estrutura da tabela dinâmica através de códigos HTML dentro do PHP, linhas 6 a 14. Nas linhas 16 a 21 ocorre inserção dos valores obtidos pela consulta do MySQL do módulo de banco de dados em vetores específicos de cada valor retornado da tabela.

Com esses vetores formados, há uma estrutura de randomização com controle de repetição, linhas 25 a 36, para que a tabela com as recomendações sejam apresentadas ao usuário.

De forma semelhante ocorre no módulo de recomendação do MPEG-7. O Algoritmo 5.6 representa o código da formação de uma lista de reprodução, recomendada ao usuário, utilizando o padrão MPEG-7. As oito listas geradas pelos oito descritores utilizados serão discutidas e analisadas na próxima seção desse projeto.

Algoritmo 5.6 Algoritmo do Módulo de Recomendação do MPEG-7

```

1: $aux = array_search(min($vetor), $vetor);
2: unset($vetor[$aux]);
3:
4: for($i=0; $i<10; $i++){
5:   $aux = array_search(min($vetor), $vetor);
6:   $resultfilho[$i] = $aux+1;
7:   unset($vetor[$aux]);
8: }
9:
10: echo "<table width='70%' border='0' align='center'>
11: <tr>
12: <td width='8%' align='left' valign='middle'><font color='red'><b>ID
13: </b></font></td>
14: <td width='90%' align='left' valign='middle'><font color='red'><b>
15: NOME DA MUSICA</b></font></td>
16: </tr>";
17:
18: for ($i=0 ;$i<10 ; $i++){
19:   $sql = "SELECT * FROM mpeg7 where id=$resultfilho[$i]";
20:   $query = mysql_query($sql);
21:   $resultado = mysql_fetch_assoc($query);
22:   $nome=$resultado['nome'];
23:   $id=$resultado['id'];
24:   echo "<tr>
25: <td align='left' valign='middle'>".$id."</td>
26: <td align='left' valign='middle'><a href='musica.php?id=$id'>
27: ".$nome."</a></td>
28: </tr>";
29: }
30: echo "</table>";

```

Através da análise desse código, é possível verificar na linha 1 e 2 que há um procedimento que busca o índice que possui o menor valor obtido pelo cálculo da distância euclidiana.

Esse primeiro valor retornado será zero e será descartado, pois é o resultado do cálculo da distância euclidiana da música que o usuário selecionou por ela mesma, pois todas as músicas são comparadas dentro do laço de repetição inclusive a própria música selecionada.

Com o comando *unset* do PHP é removido o índice do vetor que contém o menor valor especificado na função para que o próximo menor valor seja obtido.

Realizando o mesmo procedimento citado acima nas linhas 4 a 8, ocorre a criação de um vetor resultante, *resultfilho*, que recebe os índices que possuem os menores valores armazenados. Como já foi comentado, esses índices possuem relação direta com os ID's das músicas na tabela do banco de dados. Dessa forma, esse vetor se torna ordenado de forma crescente utilizando os resultados das distâncias euclidianas.

Com esse vetor resultante, seus dez primeiros índices, que contém os ID's das músicas que possuem as menores distâncias euclidianas com a música selecionada pelo usuário, são utilizados. Assim, uma tabela com todos os dez títulos musicais recomendados ao usuário é criada, linhas 18 a 30.

5.3 Resultados

Na etapa inicial de implementação desse sistema, se tornou necessário a modelagem do banco de dados com suas tabelas. Os conteúdos dessas tabelas foram inseridos conforme explicado nas seções anteriores.

Porém, como as músicas que compõem a base de dados desse sistema vêm de origens distintas, elas podem não trazer as informações corretas de seu gênero em suas etiquetas ou até mesmo, através do uso do MPEG-7, existir recomendações que não possuam relação nenhuma de gênero entre a música selecionada e as músicas recomendadas.

A Figura 5.5 apresenta uma limitação individual das etiquetas ID3 que é verificada pelas informações conflituosas obtidas através da leitura do gênero musical de músicas de um mesmo artista. Assim, fica explícito a possibilidade das etiquetas ID3 trazerem informações incorretas ou conflituosas com o verdadeiro gênero musical a qual a música pertence. Nessa Figura, a primeira coluna representa os ID's das músicas, a segunda coluna representa o título musical e a última coluna representa o gênero extraído das etiquetas ID3.

Assim, se torna necessário a criação de um mecanismo que verifique o real gênero das músicas para dar subsídios a uma análise mais completa das recomendações feitas com o uso do MPEG-7 e das Etiquetas ID3.

| | | |
|-----|---|-------------|
| 506 | Legiao Urbana - Andrea Doria | Rock & Roll |
| 507 | Legiao Urbana - Como E Que Se Diz Eu Te Amo | - |
| 508 | Legiao Urbana - Dezesseis | Hard Rock |
| 509 | Legiao Urbana - Eduardo e Monica | Rock |
| 510 | Legiao Urbana - Indios | genre |
| 511 | Legiao Urbana - Meninos e Meninas | Pop |

Figura 5.5: Limitação Individual das Etiquetas ID3

Essas mil músicas foram cedidas, para esse projeto, por diversos usuários que as obtiveram exclusivamente da internet. Essas mesmas músicas foram utilizadas para a base de dados tanto para o sistema desenvolvido utilizando individualmente o ID3 e o MPEG7 quanto para o desenvolvimento do MUSID7, visto no próximo Capítulo. Essa quantidade de músicas foi escolhida por abranger uma boa quantidade diferenciada de gêneros musicais e foram utilizadas sem sofrerem nenhuma alteração nos seus metadados. A relação completa de gêneros musicais extraídos das etiquetas ID3 das mil músicas da base de dados se encontra no Apêndice A.

Atualmente se encontram disponíveis na internet diversas comunidades e sistemas de recomendações de músicas baseados em metadados. Segundo Corthaut et al. (2008), o *MusicBrainz* é uma comunidade baseada em metadados musicais utilizada para reconhecimento de músicas. Essa comunidade traz diversas informações sobre artistas, álbuns e/ou uma determinada música.

O *MusicBrainz Picard* é um exemplo de aplicação do *MusicBrainz* que foi utilizada para identificar faixas de áudio contidas na base de dados desse projeto. A aplicação facilita a identificação das músicas ao comparar as faixas de áudio com uma base de dados online do *MusicBrainz* através de recursos de exploração e análise acústica da música. Após a identificação, essa aplicação é capaz de preencher automaticamente as etiquetas nos arquivos de áudio.

Citado por Bauer, Jansen e Cirimele (2011), como um recurso de identificação de preferência musical, o *Last.fm* é uma fonte de metadados para um sistema de recomendação de músicas onde um determinado usuário recebe recomendações de acordo com as músicas que ele ouve, ficando nos metadados a base do processo.

O *Last.fm* possui aplicativos próprios que permitem o usuário enviar para o banco de dados do seu sistema as informações presentes nas etiquetas das músicas de sua lista de reprodução. O *Scrobbler* é um recurso que permite esse tipo de conexão entre a lista de reprodução de um usuário e as informações da base de dados do *Last.fm*. Através da reprodução de uma determinada música nesse reprodutor, o usuário passa a ter acesso a informações detalhadas sobre o artista, álbum, notificações de turnê e muito mais. Além disso, é possível, dentro da própria página do *Last.fm*, realizar consultas individuais de faixas de áudio com o retorno de informações acerca dessa determinada música pesquisada.

Para Furht (2009) e Kastner et al. (2002), o *Shazam* é um sistema de identificação musical baseado em *fingerprinting* (impressão digital), ou seja, ele é capaz de identificar uma música apenas por um parte da mesma. Além de identificar uma música, o *Shazam* fornece diversas informações sobre o áudio identificado como gênero musical, artista e álbum.

Disponível para diversas plataformas, o *Shazam* pode ser instalado em um computador para realizar o reconhecimento de faixas de músicas que são reproduzidas. Esse recurso também foi utilizado para a identificação das músicas da base de dados desse projeto. Após instalado, a aplicação foi capaz de identificar músicas e fornecer diversas informações sobre elas inclusive o gênero musical.

Com o uso desses três recursos (*MusicBrainz*, *Last.fm* e *Shazam*) de forma complementar e com a obtenção do gênero real o qual pertencem as músicas que compõem a base de dados desse sistema, é possível a criação de um mecanismo que analise o gênero das músicas recomendadas tanto pelas Etiquetas ID3 quanto pelo MPEG-7 e estabeleça um grau de concordância entre o gênero real da música escolhida e os gêneros reais das músicas recomendadas.

Por exemplo, se a música selecionada pelo usuário possui um gênero real classificado como MPB e na lista de recomendações existem somente duas músicas com o gênero real também MPB, então há uma concordância de vinte por cento daquela lista de reprodução, pois o sistema recomenda apenas dez títulos.

Assim, foi criada uma tabela chamada *metadados* que contém três colunas: *ID*, *Nome* e *GeneroReal*, sendo que esta última traz as informações dos gêneros reais das músicas da

base de dados obtidos pelos mecanismos citados acima.

Para medir o desempenho entre os oito descritores utilizados do MPEG-7, além da concordância, o tempo de processamento das recomendações e a memória utilizada nos processos foram analisados. Dessa forma, é possível verificar qual dos oito descritores possui uma maior eficiência na recomendação de músicas. Todos os processamentos desse projeto foram realizados em um MacbookPro com processador Core i5 de 2.5 Ghz com 8GB de memória DDR3.

5.3.1 Análises Experimentais

Foram selecionadas cinquenta músicas de forma aleatória para que fossem geradas suas respectivas listas de recomendação utilizando o gênero da ID3 e todos os oito descritores do MPEG-7.

Após todos os processamentos envolvendo as cinquenta músicas, foram calculadas as médias aritméticas dos graus de concordância para cada metadado além do tempo de processamento e da memória utilizada para cada descritor do MPEG-7. Assim, a Figura 5.6 apresenta o gráfico das médias aritméticas entre as cinquenta músicas processadas para o fator tempo medido em segundos.

Nesse gráfico, o eixo das ordenadas representa os tempos e o eixo das abscissas representa os descritores do MPEG-7. Esse gráfico mostra o tempo da recomendação das músicas utilizando o descritor *Basis* como o melhor entre os demais descritores do MPEG-7.

A Figura 5.7 apresenta a média de porcentagem da concordância entre o gênero real da música selecionada e os gêneros reais das músicas recomendadas.

Essa concordância, como foi citado no início dessa seção, utilizou uma tabela de metadados que traz os verdadeiros gêneros das músicas utilizadas nesse projeto. Nesse gráfico, as Etiquetas ID3 apresentou o melhor nível de concordância com uma média de 33,59 por cento. Logo após, o descritor *Basis* apresentou um nível de concordância de 18,20 por cento sendo o maior grau de concordância entre os descritores do MPEG-7.

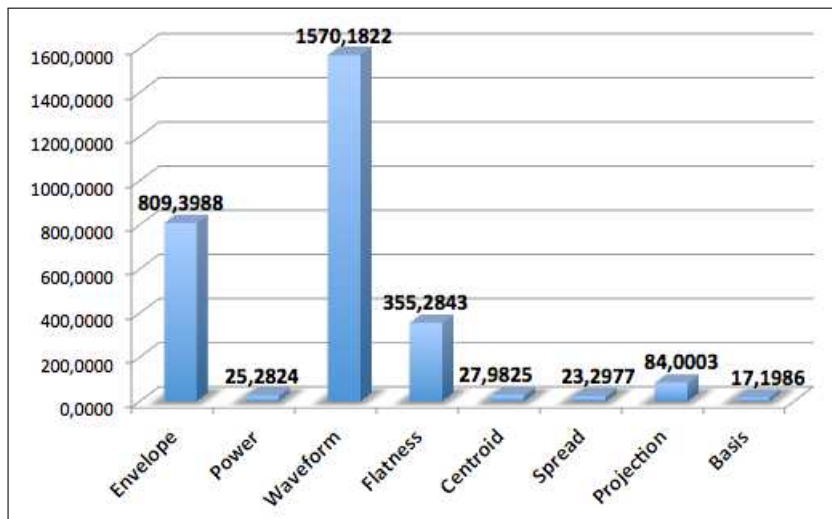


Figura 5.6: Gráfico das Médias dos Tempos de Recomendação

Devido aos procedimentos de implementação das etiquetas ID3 serem diferentes dos descritores do MPEG-7, as etiquetas ID3 não entraram em comparação de desempenho de tempo e memória, ficando eles somente para os descritores do MPEG-7.

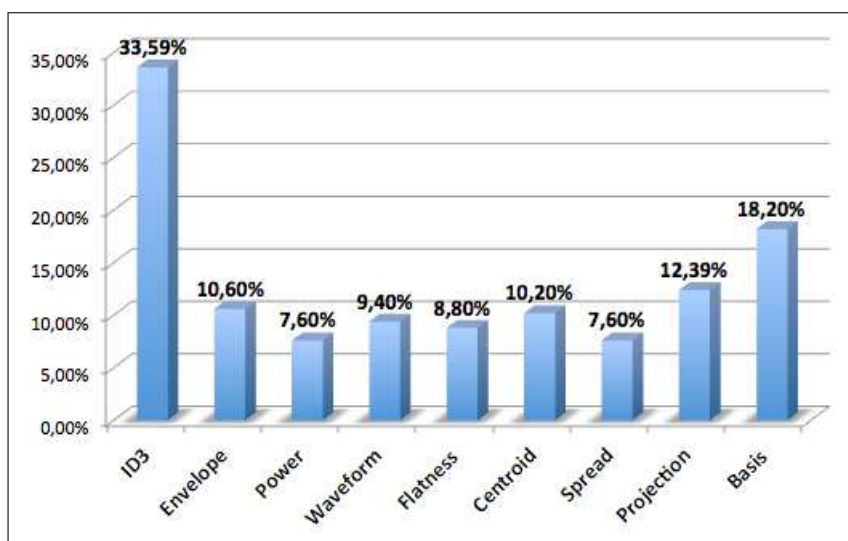


Figura 5.7: Gráfico das Médias de Porcentagem das Concordâncias nas Recomendações

A diferença das médias entre o ID3 e os descritores do MPEG-7 apresentada na Figura 5.7 pode ser atribuída ao fato de que os vetores dos descritores do MPEG-7 podem apresentar uma mudança de instancia em pouco tempo. Essa mudança pode conduzir uma determinada música de uma classificação para outra dentro dos diversos gêneros e subgêneros musicais conhecidos.

Brecheisen et al. (2006) fornecem um exemplo de hierarquia entre os gêneros e subgêneros utilizados em seu projeto e propõem técnicas de redução de dimensão vetorial para a classificação de músicas. A Figura 5.8 representa o exemplo de hierarquia de gêneros musicais.

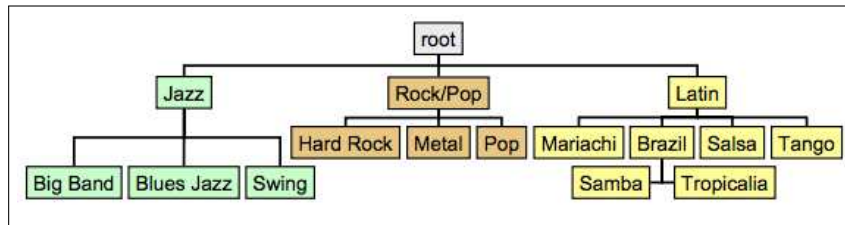


Figura 5.8: Exemplo de Hierarquia de Gêneros Musicais.

Fonte: Brecheisen et al. (2006).

Além disso, essa diferença pode ficar mais ou menos acentuada de acordo com a veracidade das informações presentes nas etiquetas ID3. Por ser de fácil manipulação, qualquer usuário pode inserir a informação que lhe seja pertinente nos espaços destinados às etiquetas ID3. Dessa forma, raramente o ID3 atingirá cem por cento de concordância e além disso, esse resultado pode ser completamente diferente caso envolva uma outra amostra de músicas.

A memória utilizada pelos descritores do MPEG-7 no processo de recomendação, representada em *KByte*, também teve sua média calculada e representada no gráfico da Figura 5.9.

Nesse gráfico, a memória utilizada pelo descritor *Basis* do MPEG-7 foi a menor memória necessária para o processo de recomendação entre os oito descritores, utilizando aproximadamente 539,7 KB. Assim, esse descritor, dentre os outros descritores do MPEG-7, foi o que utilizou menos memória durante o processo de recomendação.

O sistema de recomendação proposto neste trabalho, o qual é apresentado no próximo Capítulo, tem como objetivo um possível aumento da eficiência na recomendação de músicas baseada em gêneros musicais em relação às recomendações feitas utilizando somente o gênero das etiquetas ID3 ou somente os descritores do MPEG-7. Sendo assim, o sistema de recomendação proposto busca integrar as recomendações obtidas a partir do gênero das etiquetas ID3 com as recomendações obtidas a partir de um dos descritores do MPEG-7.

A escolha de apenas um dos descritores do MPEG-7 ocorreu após serem realizados testes

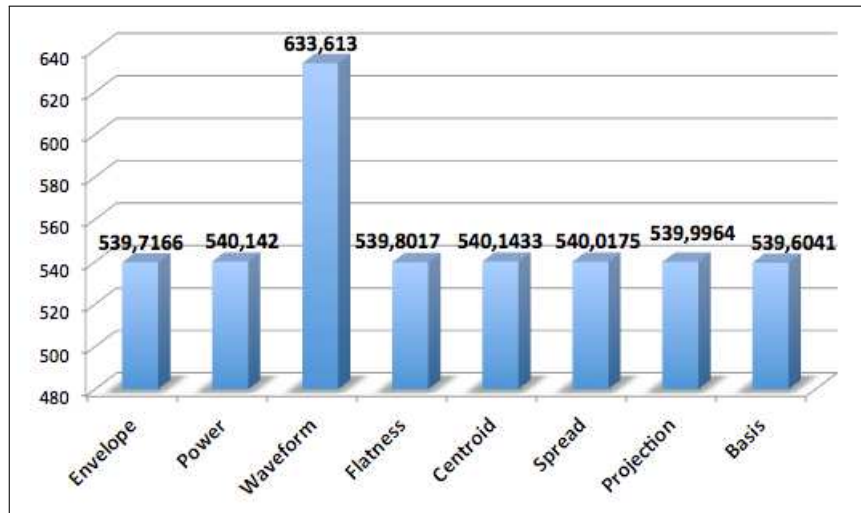


Figura 5.9: Gráfico de Médias de Memória Utilizada nas Recomendações

experimentais envolvendo sequencialmente dois ou mais descritores deste padrão com a finalidade de aumentar o grau de eficiência da recomendação, porém sem êxito.

Esse processo ocorreu, por exemplo, criando um subgrupo de músicas que possuíam os maiores graus de similaridade com a música selecionada pelo usuário, utilizando um determinado descritor. Outro subgrupo era formado do primeiro com as músicas que possuíam os maiores graus de similaridade com a música selecionada, utilizando um outro descritor. O último subgrupo formado era apresentado ao usuário com as músicas recomendadas.

Assim, analisando as três métricas que foram utilizadas nos descritores do MPEG-7 de forma individual, nota-se que o descritor *Basis* obteve um melhor resultado e, por esse motivo, foi escolhido para ser implementado no sistema de recomendação apresentado no próximo Capítulo.

Capítulo 6

MUSID7: Um Sistema para Recomendação de Músicas Integrando ID3 e MPEG-7

Neste capítulo são descritos os aspectos de um sistema de recomendação de músicas que utiliza os metadados do ID3 e do MPEG-7 de forma integrada, propondo uma possível otimização do grau de concordância nas recomendações em relação ao uso individual de cada um desses metadados.

6.1 Arquitetura

A implementação de um sistema integrado tem como objetivo diminuir as limitações do uso individual dos conjuntos de metadados ID3 e MPEG-7 ao tempo que a eficiência na recomendação seja aumentada.

Para esse propósito, foi desenvolvida uma arquitetura de um sistema que busca a interseção entre as recomendações das Etiquetas ID3 e as recomendações do descritor *Basis* do MPEG-7.

A Figura 6.1 representa a arquitetura desse sistema. Até o fluxo 6, as etapas de

processamento do MUSID7 são iguais ao sistema descrito no Capítulo anterior. A diferença se encontra a partir do módulo de recomendação.

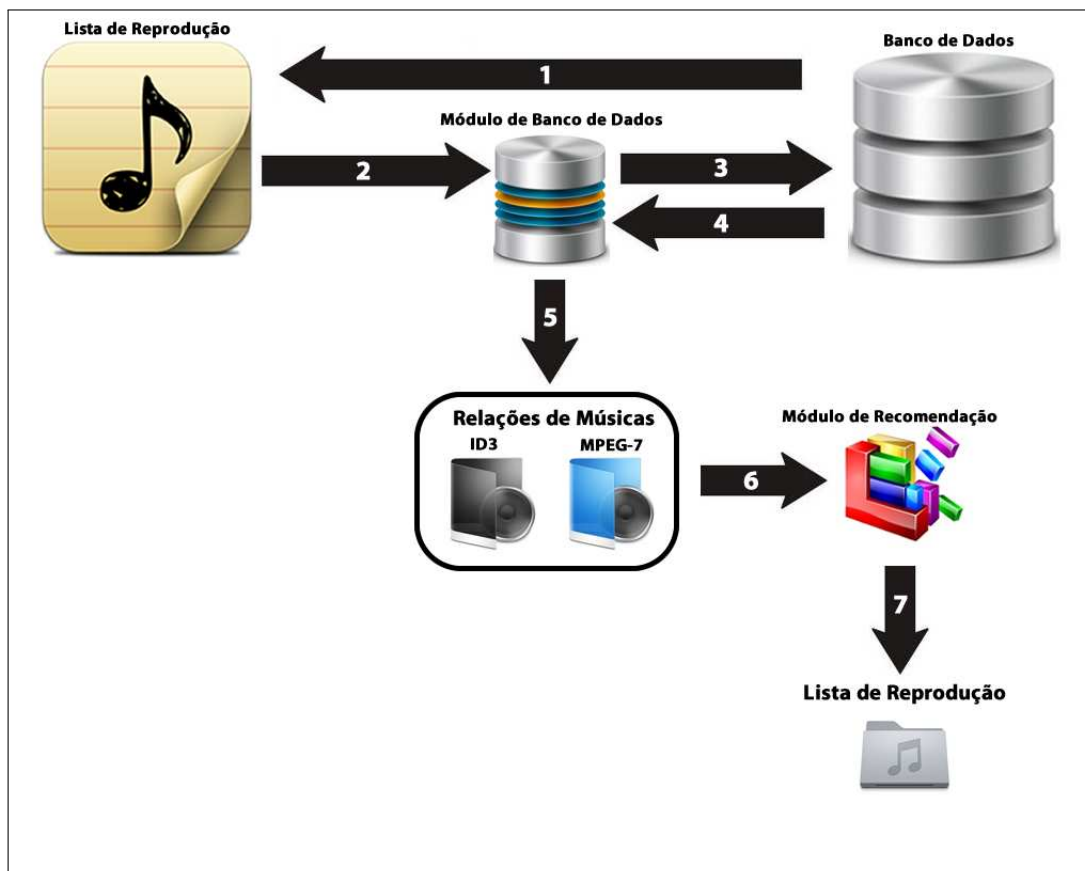


Figura 6.1: Arquitetura do MUSID7.

Após a geração das duas listas de músicas candidatas representada no fluxo 5, sendo uma gerada utilizando o ID3 e a outra o descritor *Basis* do MPEG-7, ocorre a busca pela interseção entre essas duas listas no módulo de recomendação.

A Figura 6.2 apresenta o processo de busca pela interseção. A lista gerada pelo descritor do MPEG-7 possui todas as músicas contidas na base de dados, com exceção da música que foi escolhida pelo usuário, e está ordenada através do cálculo da distância euclidiana. Essa lista é lida de forma crescente pela distância euclidiana, uma música de cada vez, e as faixas musicais que também existirem na lista de músicas do ID3 tem seus *IDs* registrados em um outro vetor.

Dessa forma, esse vetor que possui os *IDs* registrados tem todas as músicas presentes na lista do ID3, porém ordenadas pelos cálculos da distância euclidiana do descritor do MPEG-7.

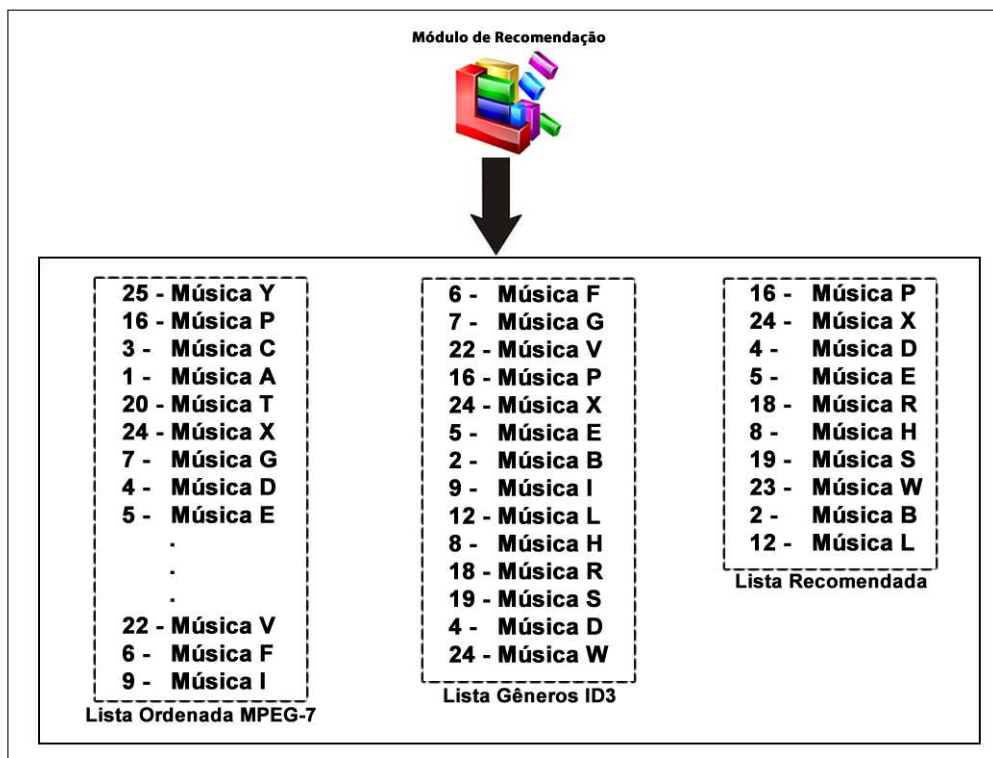


Figura 6.2: Representação da Recomendação do MUSID7.

O fluxo 7 representa justamente a saída desse vetor formado do módulo de recomendação. Esse fluxo está simbolizando a apresentação da lista de reprodução final, com a interseção das duas listas, ao usuário.

6.2 Implementação

A implementação desse sistema ocorre também com o uso do PHP e o banco de dados MySQL, semelhante ao sistema descrito no Capítulo anterior.

O Algoritmo 6.1 representa a estrutura da implementação do MUSID7. Nesse algoritmo é visto na linha 1 o procedimento de obtenção do ID da música selecionada pelo usuário. A partir desse ID, há a criação do vetor com os valores do cálculo da distância euclidiana entre a música selecionada e as demais da base de dados, linha 2, através de uma função que é detalhada através do Algoritmo 3.4 presente no Capítulo 3.

A linha 3 e 4, exclui do vetor o índice com o menor valor. Nessa primeira exclusão será

Algoritmo 6.1 Algoritmo do MUSID7

```

1:  $sid_escolha=$_GET['id'];
2:  $vetor = Descritor_SpectrumBasis($sid_escolha);
3:  $aux = array_search(min($vetor), $vetor);
4:  unset($vetor[$aux]);
5:  for($i=0; $i<999; $i++){
6:    $aux = array_search(min($vetor), $vetor);
7:    $resultMPEG7[$i] = $aux+1;
8:    unset($vetor[$aux]);
9:  }
10: $sql = "SELECT * FROM etiquetas where id=$sid_escolha";
11: $query = mysql_query($sql);
12: $resultado = mysql_fetch_assoc($query);
13: $genero = $resultado['genero'];
14: $query = mysql_query("SELECT * FROM etiquetas where genero='$genero'");
15: $num = mysql_num_rows($query);
16: $i=0;
17: while ($resultado = mysql_fetch_assoc($query)){
18:   $id = $resultado['id'];
19:   if($id <> $sid_escolha){
20:     $resultID3[$i]=$id;
21:     $i++;
22:   }
23: }
24: \\Início do Módulo de Recomendação
25: $k=0;
26: $j=0;
27: while((($j<10) && ($j<$i))){
28:   $id = $resultMPEG7[$k];
29:   if(in_array($id, $resultID3)){
30:     $final[$j] = $id;
31:     $j++;
32:   }
33:   $k++;
34: }
35: \\Fim do Módulo de Recomendação

```

excluído o índice que recebeu o valor zero, pois é o resultado da distância euclidiana entre a música escolhida e ela mesma vinda da base de dados.

Nas linhas 5 a 9 ocorre o ordenamento de um vetor pelo grau de similaridade através do resultado do cálculo da distância euclidiana. As primeiras posições desse vetor recebem os *IDs* das músicas que apresentaram os resultados menores do cálculo da distância.

Com o primeiro grupo formado através do ordenamento do cálculo da distância euclidiana, o segundo grupo de músicas é feito pelo retorno da consulta no banco de dados de quais são as faixas musicais que possuem o mesmo gênero, gravado na etiqueta ID3, da música

selecionada, linhas 10 a 15.

Com os dois grupos prontos, as linhas 25 a 34 apresentam o procedimento para gerar um vetor com a relação de músicas gerada pelo ID3 e ordenadas de acordo com a distância euclidiana do descritor do MPEG-7. Por fim, resta exibir o vetor *final*, presente na linha 30, que contém a interseção entre os dois grupos de recomendação.

6.3 Resultados

Como foi citado na seção anterior, para a implementação do MUSID7 algumas etapas foram semelhantes as etapas do sistema descrito no Capítulo 3. Dessa forma, os métodos realizados para a obtenção da média aritmética da concordância de gêneros musicais com o uso do MUSID7 foi o mesmo, usando as mesmas cinquenta músicas que foram utilizadas no cálculo da concordância do sistema de recomendação de músicas com o uso individual de metadados.

O diferencial do MUSID7 é visto no processamento das duas listas formadas com a relação de músicas candidatas a serem recomendadas, sendo uma lista gerada com o uso do ID3 e a outra lista gerada com o descritor *Basis* do MPEG-7. Esse processamento diferenciado é realizado na busca da interseção entre essas duas listas com a apresentação de dez músicas ao usuário.

Após realizar todos os procedimentos com o MUSID7 e gerar a lista de recomendação com o uso desse sistema, foram obtidos os resultados apresentados na Figura 6.3. Essa Figura apresenta o resultado da média do grau de concordância de gêneros musicais. A média foi calculada com o uso das mesmas cinquenta músicas utilizadas para as análises experimentais do Capítulo anterior.

Ao analisar o gráfico, foi verificado que o sistema de recomendação de músicas baseado em gêneros musicais que utilizou os metadados de forma individual obteve uma média de concordância menor em relação ao MUSID7. Assim, a proposta de otimizar o grau de concordância nas recomendações em relação ao uso individual do ID3 e dos descritores do MPEG-7 foi atendida através do uso de um sistema que utilizasse esses metadados de forma integrada.

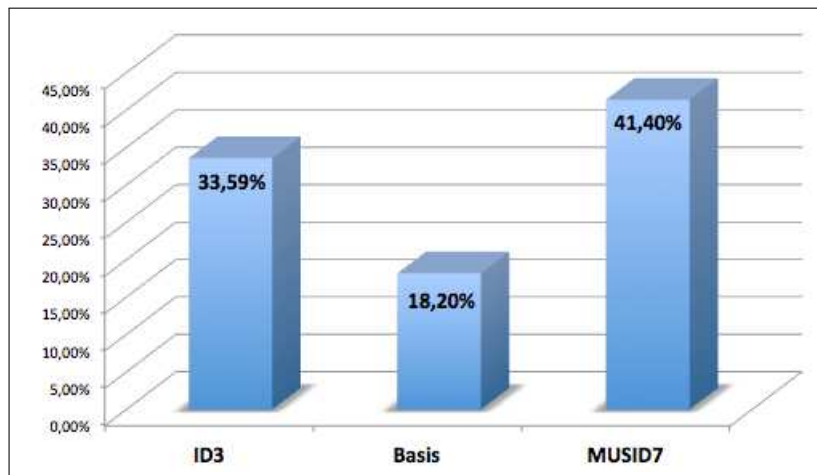


Figura 6.3: Média do Grau de Concordância de Gêneros Musicais do MUSID7

É importante salientar que os resultados aqui apresentados não podem ser tratados como absolutos, pois alterações em variáveis utilizadas nesse projeto podem implicar em uma apresentação de resultados diferentes. Uma alteração na quantidade de músicas do banco de dados e até mesmo uma alteração na quantidade de músicas utilizadas para o cálculo da média aritmética das métricas utilizadas podem ser significativas ao ponto de alterar os resultados.

Além disso, a escolha de outro grupo de mil músicas para compor a base de dados utilizada nesse projeto pode trazer um conjunto diferente de metadados, principalmente nas etiquetas ID3. Dessa forma, poderia haver modificações nos resultados das métricas utilizadas nas análises experimentais desse projeto, apesar de manter a mesma quantidade de músicas da base de dados e a mesma quantidade de músicas utilizadas para o cálculo da média do grau de concordância dos gêneros musicais.

Assim, é necessário a realização de mais experimentos com outros conjuntos de músicas para ratificar os resultados apresentados ao longo do desenvolvimento desse projeto.

Capítulo 7

Conclusões e Trabalhos Futuros

Este Capítulo apresenta as conclusões desse projeto, englobando as limitações, contribuições deste trabalho e trabalhos futuros.

7.1 Conclusões

Com o surgimento de novas tecnologias e novos recursos audiovisuais, o acesso a arquivos multimídia disponibilizado na internet tem aumentado significativamente. Porém, grande parte dos usuários não dispõem de recursos para classificar, organizar e recuperar esses recursos em um repositório. Nesse contexto surgem os metadados com a finalidade de fornecer ao usuário informações necessárias para uma classificação dos arquivos multimídia. As etiquetas ID3 e os descritores do MPEG-7 são exemplos de conjuntos de metadados que foram utilizados durante o desenvolvimento desse projeto.

O uso de etiquetas ID3 tem aumentado nos reprodutores de áudio para dar ao usuário mecanismos de classificação de músicas. Além disso, essas etiquetas têm sido utilizadas por comunidades de metadados para recomendar músicas a outros usuários que possuem semelhança entre suas preferências musicais, como por exemplo a comunidade do *Last.fm*.

Porém, essas etiquetas são de fácil manipulação e além disso, geralmente cabe ao usuário o seu preenchimento. Dessa forma, ocorre uma grande discordância entre as informações

presentes nas etiquetas e a música na qual estão inseridas.

Essa discordância foi visualizada nesse projeto ao ler as etiquetas das músicas e inseri-las na tabela do banco de dados. Foi observado nesse processo, músicas de um mesmo artista sendo classificadas em mais de dois gêneros musicais e até mesmo músicas sem possuir nenhuma informação presente nas etiquetas.

Por outro lado, o MPEG-7, em seu componente de áudio - ISO/IEC 15938-4, é um padrão de descrição de conteúdo multimídia para prover um rico conjunto de recursos capazes de resumir as características essenciais do sinal de áudio. Porém, devido a sua complexidade, muitas vezes é difícil escolher uma correta descrição para realizar um extração de uma determinada característica.

Para a utilização do padrão MPEG-7 nesse projeto foram utilizados descritores de baixo nível. Esses descritores caracterizam um determinado comportamento do sinal de áudio sem a necessidade de avaliar um conjunto de eventos e conceitos abstratos. Porém, foi observado que músicas de gêneros diferentes podem apresentar um comportamento semelhante dentro do uso de um determinado descritor no processo de recomendação.

Isso ocorre devido ao vetor de características ser gerado através da extração de comportamentos do sinal de áudio em pouco tempo de observação, por exemplo, em 3 segundos de observação é gerado um vetor com 100 posições. Assim, dentro de poucos segundos, o vetor de características podem apresentar uma grande variação em sua estrutura que pode levar a música de uma classificação para outra. Assim, uma música pode transitar na classificação entre os diversos gêneros e subgêneros musicais.

Por conta dessas limitações envolvendo o ID3 e os descritores do MPEG-7, um sistema de recomendação de músicas baseado em gêneros musicais que utilize um desses conjuntos de metadados de forma individual tem grandes chances de apresentar uma recomendação de músicas com discordância da real preferência do usuário.

Com a grande quantidade de músicas disponibilizada nos meios digitais e sem um método de classificação de músicas que seja eficiente, o usuário fica perdido dentro das distorções de classificação que poderão ocorrer perdendo tempo em tentar classificar as músicas dentro de seu interesse.

Dessa forma, um sistema de recomendação de músicas baseado em gênero musical que utilize o ID3 e o MPEG-7 de forma integrada foi proposto para superar as limitações individuais do uso de cada um desses metadados otimizando o grau de concordância nas recomendações. Esse sistema batizado de MUSID7 é capaz de proporcionar um ganho efetivo de tempo na filtragem de conteúdo musical contribuindo para uma classificação mais fiel aos reais interesses do usuário.

Para validar esse sistema foi proposto um método de análise do grau de concordância entre os gêneros musicais da música selecionada pelo usuário e as músicas recomendadas. A validação é feita ao verificar que o grau de concordância do MUSID7 na recomendação de músicas foi superior se comparado com o sistema de recomendação de músicas que utilizou individualmente o ID3 ou o MPEG-7. Como foi visto no Capítulo anterior através da Figura 4.2, houve o aumento no grau de concordância através do MUSID7 quando comparado ao uso individual do MPEG-7 e do ID3 pelo sistema descrito no Capítulo 3.

Porém, os procedimentos que envolveram a implementação dos sistemas analisados nesse projeto possuem limitações. A principal limitação é verificada no tipo de arquivo de áudio, pois somente arquivos MP3 são trabalhados nesse projeto devido a grande compatibilidade desses arquivos com os métodos de extração dos metadados utilizados. Além disso, os descritores do MPEG-7 extraíram características volumosas de poucos segundos observados sem que houvesse um estudo de técnicas de redução de dimensionalidade para abranger uma maior extração e para melhor evitar as transições entre os gêneros e subgêneros musicais.

Além disso, uma alteração no número de músicas do banco de dados e/ou no número de músicas utilizadas no cálculo da média do grau de concordância, tempo de processamento e memória utilizada podem implicar em uma modificação dos resultados. Como foi visto nos Capítulos anteriores, foram utilizadas mil músicas para compor a base de dados e cinquenta músicas utilizadas para extrair a média aritmética das métricas supracitadas, se essas músicas forem substituídas por outras podem haver modificações nos resultados obtidos, ratificando a necessidade de realização de novos experimentos para garantir os resultados obtidos nesse projeto.

Dessa forma, pode-se concluir que o MUSID7 cumpriu com o objetivo de processar dois

conjuntos de metadados, ID3 e o descritor do MPEG-7, de forma integrada fornecendo uma recomendação de músicas com um grau de concordância de gêneros musicais maior em relação ao uso individual do ID3 e do MPEG-7.

7.2 Trabalhos Futuros

A área de multimídia é muito extensa e o uso de alguns metadados ainda é muito recente. Os descritores do MPEG-7, por exemplo, possuem estudos recentes de suas aplicabilidades e dentro desse contexto alguns trabalhos futuros podem ser sugeridos.

O padrão MPEG-7, dentro do contexto de áudio, apresentam descritores de alto nível, utilizados para descrever características mais complexas do sinal. Assim, o uso de descritores de alto nível integrado ao uso de outros conjuntos de metadados podem ser utilizados para propostas de aumento da eficiência em outro método de classificação.

Além disso, existem métodos de recomendação de conteúdo que podem ser desenvolvidos e acoplados nesse projeto. Por exemplo, o desenvolvimento de um sistema híbrido que trabalhe não somente com as preferências de um usuário e sim com a sua vizinhança através de uma filtragem colaborativa.

Dentro do desenvolvimento de um sistema de recomendação de músicas levando em consideração a filtragem colaborativa, diversos recursos podem ser implementados. Nilsson, Mahoney e Sundstrom (2014) citam em seu artigo a utilização das etiquetas ID3 como recurso para classificação baseado na frequência que uma música é reproduzida por um usuário. Assim, dentro de uma vizinhança, essa classificação baseada em frequências de reprodução de músicas poderia ser utilizada para gerar as recomendações.

Dentro do uso da frequência de reprodução de uma determinada música ou até mesmo com o uso de outros fatores de análise de comportamento de sinal de áudio, podem ser utilizados métodos probabilísticos de recomendação e ainda comparar esses resultados com os resultados de recomendação de comunidades de metadados como *Last.fm* ou *Shazam*.

Além disso, como foi citado por Song, Dixon e Pearce (2012), outros fatores como estilos

de vida, regiões geográficas e idades dos usuários podem ser levados em consideração para um sistema de recomendação de músicas em substituição ao gênero musical descrito por metadados.

E também, o uso de metadados que descrevem o comportamento do sinal de áudio podem ser utilizados juntamente com informações passadas pelo usuário sobre a sua rotina ou uma determinada atividade que ele esteja pensando em desenvolver.

Por exemplo, caso o usuário informe que ele está entrando em uma sala de musculação, o sistema analisará quais músicas possuem uma amplitude maior, um ritmo mais acelerado para fazer as recomendações. Diferentemente se o usuário informar que está indo dormir, nesse caso o sistema trabalhará em cima de sinais com frequências de sinal e amplitude mais baixas para gerar a recomendação.

Referências Bibliográficas

ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, v. 17, n. 6, p. 734–749, Junho 2005.

ALLAMANCHE, E. et al. Content-based identification of audio material using mpeg-7 low level description. *2nd Annual International Symposium on Music Information Retrieval 2001*, 2001.

BARCELLOS, C. D. et al. Sistema de recomendação acadêmico para apoio a aprendizagem. *X Ciclo de Palestras Novas Tecnologias na Educação, Rio Grande do Sul*, 2007.

BAUER, J. S.; JANSEN, A.; CIRIMELE, J. Moodmusic: A method for cooperative, generative music playlist creation. In: *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology*. New York, USA: ACM, 2011. p. 85–86.

BRECHEISEN, S. et al. Hierarchical genre classification for large music collections. *Multimedia and Expo, 2006 IEEE International Conference on*, p. 1385–1388, Julho 2006.

CAZELLA, S. C.; NUNES, M. A. S. N.; REATEGUI, E. B. A ciência da opinião: Estado da arte em sistemas de recomendação. *XXV Congresso da Sociedade Brasileira de Computação, São Leopoldo*, 2005.

CORTHAUT, N. et al. Connecting the dots: Music metadata generation, schemas and applications. In: *ISMIR*. [S.l.: s.n.], 2008. p. 249–254.

DALLACOSTA, A.; SOUZA, R. L. de; TAROUCO, L. M. R. A utilização da indexação de vídeos com mpeg-7 e sua aplicação na educação. *Ciclo de Palestras Novas Tecnologias na Educação, Porto Alegre*, 2004.

EIDENBERGER, H. Distance measures for mpeg-7-based retrieval. In: *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*. New York, USA: ACM, 2003. p. 130–137.

FACHINI, A. R. Classificação de áudio baseada em extração de características e redes neurais. 2011.

FURHT, B. *Handbook of Multimedia for Digital Entertainment and Arts*. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2009. 35-36 p.

- HELLMUTH, O. et al. Advanced audio identification using mpeg-7 content description. *111th Convention Audio Engineering Society*, 2001.
- HU, Y.; OGIHARA, M. Nextone player: A music recommendation system based on user behavior. *12th International Society for Music Information Retrieval Conference. Miami.*, 2011.
- HUANG, Y.-C.; JENOR, S.-K. An audio recommendation system based on audio signature description scheme in mpeg-7 audio. In: *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*. [S.l.: s.n.], 2004. v. 1, p. 639–642.
- HUNTER, J. *Adding Multimedia to the SemanticWeb: Building and Applying an MPEG-7 Ontology*, in *Multimedia Content and the Semantic Web: Methods, Standards and Tools*. [S.l.]: John Wiley, 2005.
- ISO. *Information Technology – Multimedia Content Description Interface – Part 4: Audio*. [S.l.], 2002.
- JUNIOR, D. R. *O que são tags ID3?* 2014. Tecmundo. Disponível em: <<http://www.tecmundo.com.br/musica/3119-o-que-sao-tags-id3-.htm>>. Acesso em: 10 de Junho de 2014.
- KASTNER, T. et al. Mpeg-7 scalable robust audio fingerprinting. In: *112th Convention*. Munich, Alemanha: [s.n.], 2002.
- KIM, H.-G.; MOREAU, N.; SIKORA, T. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Chichester, UK: John Wiley, 2006.
- KOSCH, H. Mpeg-7 and multimedia database systems. *SIGMOD Record*, v. 31, n. 2, p. 34–39, Junho 2002.
- MONTEIRO, M. V. S. Padrões para definição de metadados. *Universidade Federal Fluminense*, 2007.
- NACK, F.; LINDSAY, A. T. Everything you wanted to know about mpeg-7: Part 1. *IEEE MultiMedia*, v. 6, n. 3, p. 65–77, Julho 1999.
- NASCIMENTO, L. U. *Um Padrão de Metadados para Indexação e Recuperação de Objetos Multimídia*. Dissertação (Mestrado) — Universidade do Paraná, Curitiba, 2008.
- NILSSON, M.; MAHONEY, D.; SUNDSTROM, J. *ID3v2, O Público é Informado*. 2014. Disponível em: <<http://id3.org>>. Acesso em: 10 de Outubro de 2014.
- PAMPALK, E. *Computational Models of Music Similarity and Their Application in Music Information Retrieval*. Tese (Doutorado) — Vienna University of Technology, Março 2006.
- SILVA, L. C. da; GALINDO, M. Utilização de ontologias para marcação e recuperação de segmentos de vídeo mpeg-7. *Universidade Católica do Salvador, Salvador.*, 2007.
- SONG, Y.; DIXON, S.; PEARCE, M. A survey of music recommendation systems and future perspectives. *International Symposium on Computer Music Modeling and Retrieval. London.*, 2012.

SOUZA, R. G. D. de. *Aplicando Sistemas de Recomendação em Situações Práticas*. 2014. Disponível em: <https://www.ibm.com/developerworks/br/local/data/sistemas_recomendacao/>. Acesso em: 09 de Abril de 2014.

TZANETAKIS, G.; ESSL, G.; COOK, P. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE Transactions on*, v. 10, n. 5, p. 293–302, Jul 2002.

WIRTHMANN, C. H. *Metadados*. 2003. Infobras. Disponível em: <http://www.infobras.com.br/portugues/produtos_conceito_metadados.asp>. Acesso em: 10 de Setembro de 2014.

Apêndice A

Gêneros Musicais Extraídos das Etiquetas ID3

A Tabela A.1 apresenta a relação de todos os gêneros extraídos das etiquetas ID3 das mil músicas da base de dados desse projeto. Esse gêneros foram extraídos através do processo descrito na Seção 5.2 do Capítulo 5.

| Ordem | Gênero | Quantidade |
|--------------|---------------------|-------------------|
| 1º | - | 186 |
| 2º | ?????? ?????? | 01 |
| 3º | Acoustic | 12 |
| 4º | Acústico | 1 |
| 5º | Alternative | 13 |
| 6º | Ambient Alternative | 01 |
| 7º | Arrocha | 10 |
| 8º | Axé | 24 |
| 9º | Blues | 24 |
| 10º | Brazil Rock | 01 |
| 11º | Brazilian | 03 |
| 12º | Classic Rock | 01 |
| 13º | Classical | 01 |
| 14º | Club-House | 11 |

| | | |
|-----|------------------------|-----|
| 15º | Country | 22 |
| 16º | Country/Sertaneja | 24 |
| 17º | Death Metal | 01 |
| 18º | descontecido | 23 |
| 19º | Eletronic | 01 |
| 20º | Forro | 19 |
| 21º | General Alternative | 01 |
| 22º | General Blues | 01 |
| 23º | General Jazz | 08 |
| 24º | General Rock | 02 |
| 25º | General Unclassifiable | 01 |
| 26º | genre | 104 |
| 27º | Hard Rock | 02 |
| 28º | Heavy Metal | 03 |
| 29º | HIP HOP INTERNACIONAL | 01 |
| 30º | Indie | 01 |
| 31º | Jazz | 01 |
| 32º | Latin | 29 |
| 33º | Latin Rock | 01 |
| 34º | Latina | 12 |
| 35º | Love Songs | 05 |
| 36º | Metal | 03 |
| 37º | Misc | 10 |
| 38º | Miscellaneous | 02 |
| 39º | MPB | 23 |
| 40º | Nacional | 11 |
| 41º | Other | 101 |
| 42º | Pagode | 51 |
| 43º | Pop | 19 |
| 44º | ReB | 01 |
| 45º | Raggae | 39 |
| 46º | Rock | 91 |
| 47º | Rock e Roll | 04 |

| | | |
|-----|-----------------------|----|
| 48° | Rock/Pop | 07 |
| 49° | Samba | 18 |
| 50° | Soft Rock | 01 |
| 51° | Soul | 11 |
| 52° | Soundtrack | 03 |
| 53° | Television Soundtrack | 01 |
| 54° | Unknown | 15 |
| 55° | Vocal | 12 |
| 56° | World | 10 |
| 57° | www.FabricioCds.Net | 01 |
| 58° | ZookMusicas | 15 |

Tabela A.1: Relação de Gêneros Extraídos das Etiquetas ID3

Apêndice B

Tabela Comparativa dos Trabalhos Relacionados

Os trabalhos relacionados no Capítulo 4 ofereceram informações pertinentes a recursos que foram adaptados no processo de recomendação de músicas desse projeto. A Tabela B.1 apresenta os métodos utilizados em cada trabalho relacionado e o que foi adaptado para o uso.

| Trabalho Relacionado | Método Utilizado | Adaptação de Uso |
|---|--|---|
| Classificação de Gêneros Musicais de Sinais de Áudio. | Utilização de um conjunto de recursos que representam a superfície de uma música, como instrumentação e estrutura rítmica, para classificação automática de sinais de áudio. | Utilização de recursos de representação de comportamento de sinal de áudio, como os descritores do MPEG-7. |
| Uma Pesquisa de Sistemas de Recomendação de Músicas e Perspectivas Futuras | Utilização metadados editoriais, como compositor e gênero musical, para criar perfis de recomendação. | Extração e utilização dos gêneros das etiquetas ID3. |
| Nextone Player: Um Sistema de Recomendação de Músicas Baseado no Comportamento do Usuário | Utilização de um <i>feedback</i> dado por um usuário para prever itens para uma recomendação futura além de utilizar o gênero e o ano da música como fatores de análise. | O <i>feedback</i> foi adaptado para uma lista de recomendação na qual o usuário escolhe uma música de seu interesse além de também utilizar o gênero como fator de análise. |
| Classificação de Áudio Baseada em Extração de Características e Redes Neurais | Extração e utilização de características de áudio para classificação automática de sons de instrumentos musicais. | A utilização do MPEG-7 para descrever características de músicas em um determinado intervalo de tempo. |

Tabela B.1: Adaptações dos Trabalhos Relacionados