

Análise de Ferramentas de Identificação de Músicas para Apoio à Produção Multimídia

Lucas C. S. Oliveira, Karin S. Komati, Jefferson O. Andrade, Flavio S. L. de Souza

Coordenadoria de Informática – Instituto Federal do ES (IFES)

Km 6,5 - 29164-231 - ES-010 - Manguinhos – Serra – ES – Brasil

lucascs.oliveira@gmail.com, {kkomati, [joandrade](mailto:joandrade@ifes.edu.br)}@ifes.edu.br,
flavio.lamas@gmail.com

Abstract. *This paper presents a comparison between different tools to identify music as a resource to support the creation of multimedia applications. In addition, it describes the typical operation of these systems, detailing and discussing resources and features which are compared and analyzed.*

Resumo. *Este artigo apresenta uma comparação entre diferentes ferramentas de identificação de músicas, como recurso de apoio à criação de aplicações multimídia. Além disso, descreve-se o funcionamento típico destes sistemas, detalhando e discutindo quais recursos e características serão comparadas e analisadas.*

1. Introdução

Sistemas de identificação de músicas são sistemas de recuperação automática de informação no qual a entrada é uma amostra de áudio e a saída são informações, tais como o nome da música, o nome do cantor, o nome do compositor, o nome do álbum e/ou a data de lançamento. Uma situação típica é quando uma pessoa está ouvindo uma música na rádio e gostaria de saber qual é o nome da música. Nesta situação, uma solução é o uso de um aplicativo no celular, sistema de identificação de músicas, que ao ser executado irá analisar o áudio, fará uma pesquisa na base de dados e retornará a resposta à consulta.

Outra situação comum é, durante a produção de um sistema multimídia, muitas vezes, o autor lembra-se da melodia da música, mas não se lembra de características que possibilitem a sua definição, tais como o nome do cantor, o nome da música ou trechos da letra (ou a letra está em um idioma desconhecido da pessoa). Neste caso, a melhor entrada seria cantarolar ou assobiar a parte da melodia que a pessoa se lembra para obter as informações sobre a música, esta forma de pesquisa se chama “*Query by Humming*” (QbH) [Barros, 2012], funcionalidade que alguns sistemas de identificação de música apresentam.

As pesquisas das informações, em ambas as situações descritas anteriormente, são classificadas como “Recuperação da Informação Musical por Conteúdo” [Rossato, 2011], isto é, são sistemas de pesquisas com entradas não verbais. Assim, uma questão é como proceder a escolha da ferramenta de identificação de música. Quais recursos elas devem apresentar para serem utilizadas para o apoio a produção multimídia? Com base nesse questionamento, analisamos algumas ferramentas e criticamos as suas vantagens e desvantagens.

2. Sistemas de Identificação de Música

Os sistemas de identificação de músicas funcionam, em geral, como mostrado na Figura 1. O sistema é dividido em duas partes ou subsistemas: um subsistema local que faz a aquisição e o tratamento inicial da informação e um subsistema remoto que é um servidor que faz a pesquisa da informação numa base de dados. O subsistema local pode ser executado em um dispositivo móvel (celular ou *tablet*) ou de um computador (*notebook* ou *desktop*). Num computador, é possível que se tenha interface *web* ou que a ferramenta tenha que ser instalada.

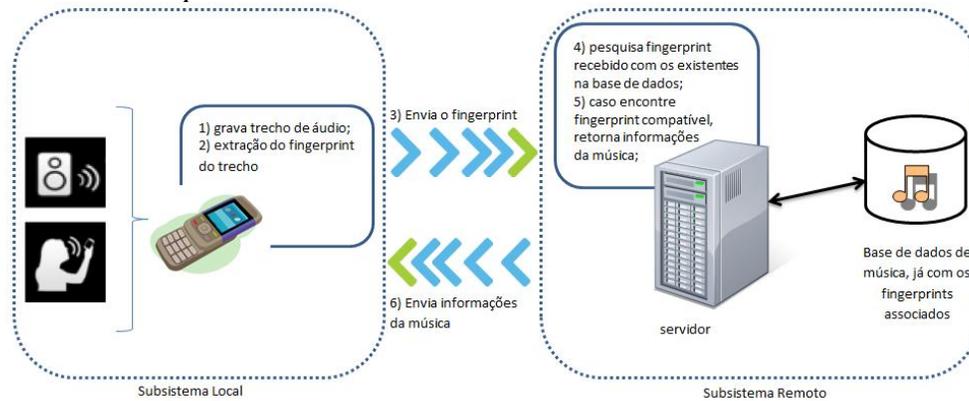


Figura 1. Funcionamento típico simplificado de um sistema de identificação de música.

A entrada da informação pode ser cantada, cantarolada ou uma música que está sendo executada numa rádio, televisão ou aparelho de som. Apenas um trecho da música é o suficiente para sua identificação. Localmente é feita a gravação do trecho e extrai-se o seu *fingerprint* associado. Assim como seres humanos que possuem impressão digital, também é possível determinar uma identificação única para a música, que é o *fingerprint* (impressão digital). No entanto, não há uma teoria única para a extração do *fingerprint*, e com isso, cada sistema adota uma forma diferente de identificação. Este tipo de busca é extremamente robusto contra distorções, comuns em microfones de baixa qualidade, e ruídos ao fundo da gravação [Typke et. al, 2005].

O *fingerprint* é enviado via rede para um subsistema remoto, onde há um servidor que irá procurar este *fingerprint* na base de dados. Não há uma base de dados única, e cada aplicativo pode ter a sua base de dados própria, onde o tamanho da base de dados é um diferencial entre as ferramentas. Outra necessidade de se ter uma base de dados própria é armazenar, além da música e de suas informações, a sua própria *fingerprint* associada, conforme dito anteriormente, pode ser diferente para cada sistema. Caso encontre uma música de *fingerprint* compatível, retorna-se as informações sobre a música, que foram encontradas na base de dados, para o subsistema local. O termo “compatível” é utilizado, pois nem sempre se usa o critério de igualdade exata, a resposta pode ser de uma música que possui o *fingerprint* mais similar.

3. Comparação e Discussão

De acordo com o que foi descrito na seção anterior, algumas características diferenciam os sistemas:

2. Tamanho da base de dados de música;
3. Tempo de carregamento no celular. Imagine a situação na qual a música está tocando numa propaganda da TV e então, uma pessoa usa o aplicativo do celular. Se o tempo de carregamento for demorado, a propaganda acabou e a pessoa perdeu a

possibilidade de saber qual era a música. Assim, é importante que o aplicativo seja leve e rápido ao ser executado;

4. Tempo de resposta. A velocidade depende muito da qualidade da conexão de rede, assim, como de quantos usuários estão acessando o servidor naquele momento. Portanto, é uma característica quantitativa que varia muito com fatores independentes da ferramenta em si;
5. Qual extensão mínima do trecho de música de entrada;
6. Qual é o tipo do subsistema local (móvel ou não) e quais são os sistemas operacionais suportadas pelo subsistema local;
7. Se o sistema é gratuito ou pago;
8. Se o sistema aceita entrada QbH;

Existem várias ferramentas disponíveis, no entanto, o mercado está bem polarizado entre duas opções: SoundHound [SoundHound Inc, 2011] e Shazam [Shazam Entertainment Ltd, 2012]. Shazam é o mais antigo, pois sua primeira versão foi lançada em 2002, enquanto o SoundHound foi lançado em 2009. Apresentamos as características selecionadas para comparação e a subsequente análise de cada ferramenta na Tabela 1, de acordo com os trabalhos de [Purdy, 2011; Chukumba, 2012; Mellon, 2012; Temporale, 2012].

Infelizmente, a informação do tamanho da base de dados do SoundHound não se encontra disponível nas páginas do produto. Em [Purdy, 2011], o autor cita que o Shazam possui uma base de dados maior que o SoundHound, mas não especifica o tamanho. O tamanho da base de dados é um fator importante, pois indica a vastidão do repertório de músicas que podem ser reconhecidas pelo sistema. A acurácia do resultado é a resposta correta do sistema, isto é, se o sistema realmente encontrou a música correta. No entanto, observamos que a acurácia depende muito da base de testes de música. Como os sistemas reconhecem apenas as músicas que se encontram na base de dados é possível que um conjunto de músicas de testes seja melhor em um sistema do que no outro, e ao trocar o as músicas do conjunto de testes a acurácia o resultado seja oposto. Muito embora, deve-se lembrar de que o tamanho da base de dados é uma característica dinâmica e pode ser aumentado rapidamente.

Tabela 1. Comparação das Características dos sistemas Shazam e SoundHound.

Nome da ferramenta	Shazam	SoundHound
Tamanho da base de dados	16 milhões	Não informado
Tempo de carregamento no celular	>7 segundos	8 segundos
Tempo de resposta média	12 segundos [Temporale, 2012]	21 segundos [Temporale, 2012] 1, 5 a 2x mais rápido[Mellon, 2012]
Trecho de áudio de entrada	2-10 segundos	5-8 segundos
Gratuito ou pago	* Free (5 músicas pesquisada/mês) * \$4.99 para Android (Shazam Encore)	* Free * \$4.99 para Android
Tipo do subsistema local/ Sistema Operacional (SO)	Celular (SO Android)	Web (Midomi) e celular (SO Android e iPhone)
QbH	Não	Sim

No teste feito entre as duas ferramentas com o mesmo celular, o Shazam foi mais rápido no tempo de carregamento do aplicativo, mas houve uma discordância quando ao tempo de resposta, em um dos trabalhos o Shazam foi melhor e no outro trabalho, o SoundHound foi melhor. O tempo do trecho de entrada depende, na prática, muito mais de qual música específica está sendo pesquisada.

Ambas as ferramentas possuem versão grátis, entretanto, o Shazam apresenta uma restrição de 5 (cinco) músicas pesquisadas por mês ao passo que o SoundHound

não apresenta esta limitação. O preço da versão paga é exatamente o mesmo valor. O SoundHound tem mais variações de ambientes em sua versão cliente, e o grande diferencial é que aceita o QbH e o Shazam não.

4. Considerações Finais

É muito importante para uma aplicação multimídia a escolha do áudio para que o processo ensino-aprendizagem seja a mais proveitosa possível. No entanto, durante a construção da aplicação, o autor pode necessitar de uma ferramenta de apoio para a identificação de músicas das quais ele quer utilizar, mas que não se lembra das suas informações associadas, nem mesmo para efetuar a compra ou o download.

Analisamos e comparamos recursos e características das duas ferramentas mais populares do mercado. Ao final, considera-se que o SoundHound é melhor, principalmente porque a sua versão de graça não possui limitações e a pesquisa do tipo QbH é essencial para o processo de apoio à autoria. Características consideradas mais importantes que sua velocidade de resposta ou carregamento.

Referências

- Barros, C. M. (2012) **Representação da informação musical: subsídios para recuperação da informação em registros sonoros e partituras no contexto educacional e de pesquisa**. 150 p. Dissertação (Mestrado em Ciência da Informação). UFSC, Florianópolis.
- Chukumba, S. (2012). I've been hard on Shazam. But SoundHound is better. Disponível em: <<http://stephenchukumba.wordpress.com/2012/02/16/ive-been-hard-on-shazam-but-soundhound-is-better/>>. Acesso em: <julho de 2012>.
- Mellon, M. (2012). **Now We Know What This Song Is. SoundHound vs Shazam**. Disponível em: <<http://www.softwarewithstyle.com/magazine/10>>. Acesso em: <julho de 2012>.
- Purdy, K. (2011) Shazam vs. SoundHound: Battle of the Mobile Song ID Services. Disponível em: <<http://lifehacker.com/5757214/shazam-vs-soundhound-battle-of-the-mobile-song-id-services>>. Acesso em: <julho de 2012>.
- Rossato, O. P. (2011) **Representação e recuperação de música em sistemas digitais: estudo exploratório**. São Paulo. 59 p. (TCC em Biblioteconomia). USP, São Paulo, SP.
- Shazam Entertainment Ltd. (2012). Shazam. Disponível em: <www.shazam.com>. Acesso em: <julho de 2012>.
- SoundHound Inc. (2011) SoundHound. Disponível em: <www.soundhound.com>. Acesso em: <julho de 2012>.
- Temporale, M. (2012) **The Music Recognition Battle: Shazam vs. Sound Hound**. Disponível em: <<http://www.mobilejaw.com/reviews/2012/06/the-music-recognition-battle-shazam-vs-sound-hound>>. Acesso em: julho de 2012.
- Typke, R., Wiering, F., Veltkamp, R. C. (2005) A survey of music information retrieval systems. In: International Symposium on Music Information Retrieval. **Anais**. pp. 153-160.