



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXVI SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2022

Desenvolvimento de ferramentas de *web scraping* para coletar informações musicais

Thiago Cardozo de Jesus¹; Bruno Westermann²

1. Bolsista PIBIC/CNPq, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: cardozothiago07@gmail.com
2. Orientador, Departamento de Letras e Arte, Universidade Estadual de Feira de Santana, e-mail: bruno.westermann@uefs.br

PALAVRAS-CHAVE: Streaming Musical; Música e Cultura Digital; Trama Virtual .

INTRODUÇÃO

Este trabalho apresenta os resultados de um plano de Iniciação Científica (IC) realizado junto à Universidade Estadual de Feira de Santana, e financiada pela Fundação de Amparo à Pesquisa do Estado da Bahia. O plano de trabalho foi desenvolvido como ação de pesquisa vinculada ao projeto "Investigações sobre música e plataformas digitais", que busca compreender como as plataformas de streaming de música têm influenciado e são influenciadas pelas práticas musicais no Brasil.

A fundamentação teórica principal do projeto de pesquisa é a teoria ator-rede (LATOURETTE, 2012; 2019) e suas associações com os campos da comunicação (LEMOES, 2013; PEREIRA DE SÁ, 2013, 2021) e da música (WESTERMANN, 2017; 2022). Desdobramento da teoria ator-rede, a cartografia de controvérsia é a concepção metodológica desta pesquisa, associada a técnicas de pesquisa para a internet (FRAGOSO, RECUERO, AMARAL; 2012) e estudos de plataforma (D'ANDRÉA, 2018; 2020).

Baseado neste lastro teórico-metodológico, este plano de trabalho de IC foi elaborado com o objetivo geral de desenvolver ferramentas de *web scraping* para coleta de dados das músicas mais ouvidas no portal Trama Virtual, a partir de seus registros no banco de dados Wayback Machine. Seus objetivos específicos foram (a) definir bibliotecas da linguagem Python coerentes com as operações a serem realizadas; (b) desenvolver ferramentas de *web scraping* para realizar as coletas de dados; (c) coletar informações do portal Trama Virtual no banco de dados Wayback Machine; (d) consolidar as informações em uma base de dados organizada e exibi-las em uma interface interativa.

A seguir, descreveremos o processo de realização da pesquisa: o desenvolvimento da ferramenta de *web scraping*, e as técnicas utilizadas para a organização dos dados. Por último, apresentaremos a análise dos dados e discussão. Os resultados desta pesquisa podem ser vistos de maneira interativa nesta ferramenta de linha do tempo, acessível em: <https://bit.ly/44IZnOD>.

METODOLOGIA

A extração dos dados foi feita através de *scripts* em na linguagem de programação *python*, como, durante o período que esteve ativa, a Trama Virtual possuiu, 3 grandes versões, foi necessário construir 3 *scripts* para coletar as informações presentes no portal. Nesse sentido cada *scripts* procura *tags* HTML específicas em cada versão do site.

Após a coleta, os dados que foram obtidos foram dispostos em planilhas, separadas por anos, contendo as informações de Música, Artista, Data do ranking e Horário. Como existiam capturas no Wayback Machine da trama virtual do mesmo dia, com a diferença apenas de horário, foi feito uma limpeza de dados, na qual consideramos que, para entradas no mesmo dia, seria considerado apenas a ultima entrada, baseado no horário.

Após a limpeza dos dados, foi necessário limitar o objeto de estudo, pois tínhamos dois tipos de dados, os provenientes do top 5 e os provenientes do top 100. Como não foi possível determinar se os dois rankings possuíam o mesmo critério de avaliação, optou-se por analisar os dados provenientes do top 5 (2005 até 2013), já que estes dados abrangem um maior período de tempo, em relação aos do top 100 (2005 até 2009). E, a fim de simplificar o armazenamento e otimizar o processo de análise, os dados coletados pelos *scripts*, antes em tabelas separadas, foram unidas numa única tabela.

A análise dos dados foi feita com base na construção de um escore, uma medida criada para verificar a incidência de músicas em um período determinado de tempo. Esse escore foi calculado utilizando a quantidade de vezes que determinada música aparece no ranking (representado pela letra Q) e multiplicando por um peso (P), que diz respeito à sua colocação no ranking. Dessa maneira, o escore (S) de uma música era resultado da soma das multiplicações entre a quantidade de vezes que ela apareceu no ranking (Q) e seu respectivo peso (P), de acordo com sua colocação.

Vale ressaltar que o escore para cada música foi feito com base nas ocorrências dela dentro do período de um ano, ou seja, uma mesma música pode ter scores diferentes em anos diferentes.

RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)

A partir dos dados coletados, é possível verificar a presença majoritária de bandas de rock entre os anos de 2005 e 2008, período sobre o qual foram coletados os primeiros dados desta pesquisa sobre a Trama Virtual. No ranking do ano de 2005, por exemplo, músicas da banda Fresno aparecem em 3 das 10 posições do ranking. Além dela, NX Zero, Treslados e Dossiê também estão presentes no ranking, todas bandas de rock. A exceção, neste caso, é a artista Noemia Duque, que aparece no ranking de 2005 em terceiro lugar, e que pode ser classificada como uma cantora de afoxé eletrônico, segundo informações do seu perfil no SoundCloud.

No ano de 2006 o ranking é composto completamente por bandas de rock. As bandas Fresno e NX Zero são as mais presentes, com três e duas músicas, respectivamente. Em contrapartida, no ano de 2007, a banda Fresno aparece apenas com uma música no ranking, que aparece em terceiro lugar; e NX Zero não aparece neste ano. O rock se mantém como único gênero entre as dez músicas mais ouvidas em 2007, tendo as bandas Treslados e Dance of Days como mais presentes, com duas músicas cada.

Em 2008, o rock permanece como gênero único, mas com bandas diferentes daquelas que formavam os rankings dos anos anteriores. Além disso, houve uma menor diversidade de bandas, se comparado com os anos anteriores. Keepers e Fused são os grupos com maior recorrência entre as 10 músicas mais baixadas, com quatro e três músicas, respectivamente. Em 2009 a ausência de diversidade de bandas continua. A banda Keepers neste ano aparece com 6 músicas no ranking.

A partir do ano de 2010 é possível verificar a recorrência de outros gêneros musicais entre as músicas mais baixadas na Trama Virtual. Este é o primeiro ano, desde 2005, que o top 10 não é formado exclusivamente por artistas de rock. O cantor Galldino aparece no ranking, em oitavo lugar, com a música Amorticínio - a canção. As demais músicas são do gênero rock, porém este ano possui uma maior diversidade de artistas. As bandas Plug3 e Dimprestables são as mais recorrentes, com duas músicas cada.

A partir disso, é possível verificar que, até 2010, existe pouca variedade de grupos musicais nos rankings. Torna-se recorrente, a partir deste momento, que uma mesma banda apareça com 3, 4 ou até 6 músicas no ranking de um mesmo ano. Entretanto, as bandas que aparecem em um determinado ano, dificilmente aparecem no próximo. Merece destaque também que a quase totalidade dos grupos musicais que aparecem nos rankings da Trama Virtual são artistas cuja carreira se desenvolveu exclusivamente em cenas musicais independentes e/ou alternativas. Há exceções a esta regra, como o caso dos grupos NX Zero e Fresno, mas estes estão presentes apenas nos rankings de 2005 e 2006. Caberia um aprofundamento na investigação destes e outros casos, para compreendermos com mais detalhes a relação entre a presença de artistas na Trama Virtual e outros movimentos dentro da indústria fonográfica, como os eventuais contratos assinados com gravadoras, por exemplo.

Nos dois anos seguintes, 2011 e 2012, o rap passa a frequentar este ranking. O rapper Lê BMX aparece na primeira colocação em 2011, com a música “Vida de BMX”; já Nitro Di, com a música “Mais um vencedor”, é o primeiro colocado de 2012. Nestes anos houve uma maior diversidade de gêneros musicais, quando comparado com os anos anteriores, podendo ser vistas músicas de rock, rap e pop. Além disso, bandas compostas majoritariamente por mulheres também aparecem nas colocações do ranking, como a banda de rock Uh La La e a iniciativa musical Mulheres Periféricas Cantam.

Relacionando os dados deste trabalho com aqueles publicados por AUTOR DO TEXTO (ANO), inseridos no mesmo projeto de pesquisa, é possível verificar que as músicas que predominam na Trama Virtual se diferem das músicas das demais plataformas. Um exemplo disso são os anos de 2012 e 2013, quando YouTube e Trama Virtual coexistiram, e os rankings dos dois serviços são completamente diferentes. De um lado, na Trama Virtual, o rock é o gênero predominante e o rap aparece como tendência secundária. No Youtube, esses gêneros dificilmente aparecem nos seus rankings, cujos gêneros predominantes são o funk e o sertanejo, gêneros esses que não aparecem em nenhum ano da Trama Virtual.

Plataformas posteriores à Trama Virtual, como Spotify e Deezer, também possuem a mesma tendência do Youtube, com exceção dos anos iniciais de ambas, quando é possível verificar presença majoritária de bandas internacionais. À medida em que essas plataformas se consolidam no Brasil, nos seus ranking começa a predominar os mesmos gêneros do Youtube, o sertanejo e o funk (AUTOR DO TRABALHO, ANO, página).

CONSIDERAÇÕES FINAIS

Acreditamos que este trabalho apresenta informações relevantes sobre o processo de digitalização da música no Brasil. A partir da associação dessas

informações com outras ações de pesquisa deste projeto, já é possível perceber que existem características importantes na Trama Virtual que são bastante distintas daquelas presentes em outras plataformas de streaming de música posteriores. Em trabalhos futuros, e com mais dados de pesquisa produzidos, acreditamos que será possível mapear outras características do processo de plataformização da música no Brasil.

REFERÊNCIAS

CUNHA, S.S; LIMA, P.R.L. 2006. Influência do tipo de reforço no comportamento à flexão de painéis laminados. In: XI Seminário de Iniciação Científica da UEFS, Feira de Santana, p.21-22.

HENNIG, W. 1981. Insect phylogeny. Chichester, John Wiley, 514p.

HERWIN, T.L.; J.C. SCOTT. 1980. Seasonal and size patterns, trophic structure, and richness of Coleoptera in the tropical arboreal ecosystem: the fauna of the tree *Lucea seemannii* Triana and Planch in the Canal Zone of Panama. *Coleopt. Bull.* 34(3): 305-322.

HULL, D.L. 1974. Darwinism and historiography. In: T.F. GLICK (ed.), *The Comparative reception of Darwinism*, pp. 388-402. Austin, Univ. Texas.

LIMA, P.R.L. 2004. Análise teórica e experimental de compósitos reforçados com fibras de sisal. Universidade Federal do Rio de Janeiro, Tese.

MORI, S.A., B.M. BOOM; G.T. PRANCE. 1981. Distribution patterns and conservation of eastern Brazilian coastal tree species. *Brittonia* 33 (2): 233-245.

POLHILL, R.M.; P.H. RAVEN (eds.) 1981. *Advances em Legume Systematics*. London, Royal Botanic Gardens Kew, 1049 p.

PUNT, W., S. BLACKMORE, S. NILSSON; A. LE THOMAS. 1999 [online]. Glossary of pollen and spore terminology. Homepage: <http://www.bio.uu.nl/~palaeo/glossary/glos-int.htm>

QUATE, L.W. 1965. A taxonomic study of Philippine Psychodidae. *Pacif. Ins.* 7(4): 815-902.

SILVEIRA, L.T. 1991. Revisão taxonômica do gênero *Periandra* Mart. ex Benth. Univ. Estadual de Campinas, MSc diss.

WESTERMANN, Bruno. As coisas e o ensino de violão: Relação entre tecnologias digitais e características do ensino do instrumento no contexto da educação a distância. 227 f. 2017. Tese (Doutorado) – Escola de Música, Programa de Pós-Graduação em Música, Universidade Federal da Bahia, Salvador, 2017.