

ANÁLISE DE SENTIMENTO APLICADA AO ESTUDO DE LETRAS DE MÚSICA

Renato Rocha Souza*
Lígia Maria Arruda Café**

RESUMO

A música vista como objeto informacional é um tema complexo quando aplicado ao campo da organização do conhecimento (OC), principalmente quando se almeja modelar um domínio de forma automatizada. Uma das técnicas que vem despontando como promissora neste setor é a Análise de Sentimento (AS) ou Mineração de Opinião; técnica derivada da inteligência artificial para identificar opiniões e emoções em textos, avaliando-as como positivas ou negativas por meio do processamento automático da linguagem natural. Esta pesquisa registra a experiência com letras de música. Após a extração de letras em várias línguas, em uma fonte específica, por meio de *crawlers*, tratou-se morfossintaticamente o léxico de letras de música em português do Brasil. Em seguida, aplicou-se a técnica de AS para classificar o corpus segundo a dimensão emocional, o que resultou em um banco de dados de 3.2 GB, com 1616 letras em 38 línguas enriquecido com informações de marcação morfossintática e polaridade emocional para o português. Esta experiência demonstrou que o aspecto afetivo pode ser medido pelo processamento automático da linguagem, especificamente via Análise de Sentimento. Os resultados levam a concluir pela efetividade da AS para letras de música, atestando sua aplicabilidade e contribuição para o avanço de pesquisas no campo a OC neste domínio.

Palavra-chave: Letras de música. Análise de sentimento. Processamento automático da linguagem natural.

* Doutor em Ciência da Informação pela Universidade Federal de Minas Gerais, Brasil. Professor e Pesquisador da Escola de Matemática Aplicada da Fundação Getúlio Vargas, Brasil. Pesquisador da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Brasil. Bolsista de produtividade do CNPq. E-mail: rsouza.fgv@gmail.com.

** Doutora em Linguística pela Universidade Laval, Canadá. Docente permanente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Brasil. Bolsista de produtividade do CNPq. E-mail: ligia.cafe@ufsc.br.

I INTRODUÇÃO

Análise de sentimento (AS) ou Mineração de Opinião é uma técnica derivada da linguística e da inteligência artificial utilizada para identificar opiniões e emoções em documentos textuais, avaliando-os como positivas ou negativas em relação a algum contexto ou objeto aludido, utilizando para isso o processamento automático da linguagem natural (PLN). Segundo Benvenuto, Ribeiro e Araújo (2015, p. 3),

polaridade representa o grau de positividade e negatividade de um texto [...] Alguns métodos tratam a polaridade como um resultado discreto binário

(positivo ou negativo) ou ternário (positivo, negativo ou neutro) [...]. Força do sentimento representa a intensidade de um sentimento ou da polaridade sendo também uma forma de saída de alguns métodos. Normalmente é um ponto flutuante entre (-1 e 1) [...]. Emoção indica um sentimento específico presente em uma mensagem (ex.: raiva, surpresa, felicidade, etc.).

As pesquisas partem do pressuposto de que a orientação semântica (ou polaridade) determina restrições de uso da palavra na língua e mostram, em que medida, o aspecto afetivo, que compõe o significado da palavra, pode estar sendo modificado, se distanciando do que é estabelecido pela norma linguística vigente.

Os campos de aplicação são os mais variados. Café e Souza (2017), em um extenso levantamento da literatura, retratam o uso desta técnica no domínio da organização do conhecimento, revelando áreas de aplicação, instituições, países, principais autores e periódicos envolvidos no registro da produção científica internacional. Os autores concluem que AS, ainda que se caracterize como uma novidade para o campo da OC, vem tomando espaço e sendo adotada nas principais tarefas de classificação de documentos. Wilson, Wiebe e Hoffmann (2005) apresentam uma experiência interessante no campo da análise de sentimento, indo além dos estudos que se limitam a somente aplicar a um corpus um léxico previamente classificado em polaridade positiva e negativa. Estes autores, considerando o funcionamento semântico das expressões no contexto linguístico em que estão inseridas, mostram que uma mesma palavra pode adquirir sentidos emocionais diferenciados dependendo da situação comunicativa em que são empregadas, constituindo uma contribuição inovadora no domínio dos experimentos voltados para a identificação automatizada de grandes conjuntos de expressões emocionais.

Aplicando também a análise de sentimento de expressões no nível da frase, os autores Jain e Nemade (2010) focam os fóruns online de domínio governamental, comercial e político com o intuito de apoiar analistas de informação na determinação de atitudes e sentimentos de eleitores e clientes.

Outra investigação relevante é a de Hatzivassiloglou e McKeown (1997) que foca a orientação semântica (positiva ou negativa) de adjetivos. Como resultado, os autores propõem um método interessante baseado no processamento da linguagem natural para identificação de categorias afetivas derivadas da análise deste tipo de classe gramatical.

No Brasil, registram-se pesquisas nesta área do tratamento da dimensão emocional por meio da linguística computacional. Destacam-se os experimentos voltados para a medição do estado de ânimo de alunos em cursos de Educação à distância. Adotando o ferramental da linguística computacional, Rigo et al. (2013) estudam a emoção com base nas expressões registradas em textos durante interações entre integrantes de Ambientes Virtuais de

Aprendizagem (AVEA). Em Balage Filho, Pardo e Aluísio (2013), é apresentada uma avaliação da versão em língua portuguesa (variante brasileira) do dicionário do Linguistic Inquiry and Word Count (LIWC). Os autores registram outros léxicos de sentimentos como o Opinion Lexicon (SOUZA et al., 2011) e o Sentilex (SILVA et al., 2012) construídos para o tratamento automático das emoções.

Citamos igualmente a pesquisa de Pang e Lee (2008) que cobre as diversas técnicas, abordagens e aplicações no campo da análise de sentimentos e mineração de opiniões na perspectiva da inteligência artificial. No que se refere às técnicas especificamente, Benvenuto, Ribeiro e Araújo (2015, p.5) relatam que os

métodos atuais de detecção de sentimentos em sentenças podem ser divididos em duas classes: os baseados em aprendizado de máquina e os métodos léxicos". O primeiro classificado como método supervisionado "é embasado nos conceitos de aprendizagem de máquina partindo da definição de características que permitem distinguir entre sentenças com diferentes sentimentos, treinamento de um modelo com sentenças previamente rotuladas e utilização do modelo de forma que ele seja capaz de identificar o sentimento em sentenças até então desconhecidas. A segunda abordagem não conta com treinamento de modelos de aprendizado de máquina e, em geral, são baseadas em tratamentos léxicos de sentimentos que envolvem o cálculo da polaridade de um texto a partir de orientação semântica das palavras contidas neste texto.

Em nossa pesquisa, tendo por objeto a informação musical, adotamos o método não-supervisionado denominado abordagem lexical, que auxiliou a classificação das letras de música (sem considerar a melodia, somente o léxico) quanto a polaridade emocional.

A música pode materializar-se de diversas formas: partituras, registros sonoros, letras de música entre outras estruturas. Há autores que acreditam serem as letras uma importante fonte de representação sobre música em sua dimensão emocional. Neste campo, Oudenne e Chasins (s.d., p.1), baseados na ideia de que "Os seres humanos coletam informações, como gênero, humor e temática de uma música

simplesmente lendo sua letra¹, examinaram, por meio do processamento da linguagem natural, 420 letras de músicas classificando-as como emocionalmente positivas ou negativas, e mostraram a complexidade da tarefa e as dificuldades relacionadas a sua automatização. Os autores determinam três maiores obstáculos no tratamento do sentimento por meio das letras de música:

1) Canções podem conter uma série de trechos negativos em suas letras, mas terminam com uma mensagem positiva, ou vice versa [...]; 2) Canções podem não conter nenhuma das pistas subjetivas capturadas em um léxico generalista, e ainda assim, expressar sentimentos positivos ou negativos 3) Canções podem expressar emoções negativas a respeito de fatos positivos, e vice versa². (OUDENNE; CHASINS. s.d., p.3).

A nosso ver, estas constatações ocorrem porque a língua não é somente a expressão do pensamento; ela é fruto de uma tomada de decisão no processo estruturado e cooperativo de comunicação, no qual ao menos dois falantes (o compositor e o ouvinte) interagem e defendem pontos de vistas. Nossa investigação vai nesta direção e oferece à comunidade científica um banco de dados com a marcação sintática do léxico das letras de música classificadas segundo a emoção. A intenção é tornar disponível aos pesquisadores uma rica fonte estruturada de dados para ser utilizada em pesquisas que procurem evidenciar os fenômenos linguísticos e emocionais relativos a informação musical.

2 CONSTRUÇÃO DE UM BANCO DE DADOS DE INFORMAÇÃO MUSICAL

As letras das canções transmitem mensagens de caráter afetivo que influenciam e são influenciadas por valores sociais e políticos de uma época. Por meio de seus conteúdos linguísticos e melódicos, registram ideias e

pensamentos de uma sociedade. Música é história e recordação. Do ponto de vista linguístico, observar os aspectos morfológicos, sintáticos e semânticos das letras nos ajudam a identificar estas visões de mundo carregadas de emoção e beleza. Percebe-se, portanto, que o sentimento é um elemento fundamental para os estudos sobre o funcionamento linguístico das letras de música. Com isto em mente, iniciou-se esta pesquisa cujo objetivo foi tratar morfossintaticamente o léxico das letras de música e classificá-las segundo a dimensão emocional.

A pesquisa iniciou com a análise e seleção das fontes de informação disponíveis na Internet que abarcavam volumes razoáveis de letras de música e dados relacionados. Foram examinadas as fontes Vagalume³, Letras⁴ e Cifraclub⁵, sendo este último mais voltado para o oferecimento de cifras e partituras das canções disponibilizadas. Dentre as opções, escolheu-se a coleta de informações de letras de música da plataforma Vagalume por esta oferecer um grande volume de canções disponíveis e por se aproximar mais do escopo da pesquisa. O rodapé do referido sítio da internet apresenta advertência⁶ quanto à reprodução da informação em outros meios, e sobre o uso educacional a que se propõe. Entendemos que a presente pesquisa se enquadra no contexto de pesquisa científica, sendo que não há intenção de reprodução do conteúdo em nenhum outro meio, para nenhum outro fim que não seja o de fomentar o conhecimento científico, nas áreas denominadas “Linguística de Corpus” e “Ciência da Informação”.

O quadro complexo do universo investigado, caracterizado pela tipologia e natureza diversificada dos dados linguísticos, nos levou a decidir pela construção de um banco de dados que constituísse em uma fonte de dados para futuras pesquisas. Tendo esta finalidade mais ampla em vista, foram desenvolvidas duas atividades para as quais contamos com o apoio da expertise do Núcleo de Análise e Modelagem de Dados (NAMD) da Escola de Matemática Aplicada da Fundação Getúlio Vargas. Iniciou-se, assim, a construção

1 Humans gather meta-information such as the genre, sentiment, and theme of a song simply by reading its lyrics

2 1) Songs can contain a series of negative lyrics but end on an uplifting, positive note, or vice versa [...]; 2) Songs may not contain any of the subjectivity clues in a general subjectivity lexicon, yet express positive or negative emotions [...]; 3) Songs can express positive emotions about negative things, and vice-versa

3 <https://www.vagalume.com.br/>

4 <https://www.letras.mus.br/>

5 <https://www.cifraclub.com.br/>

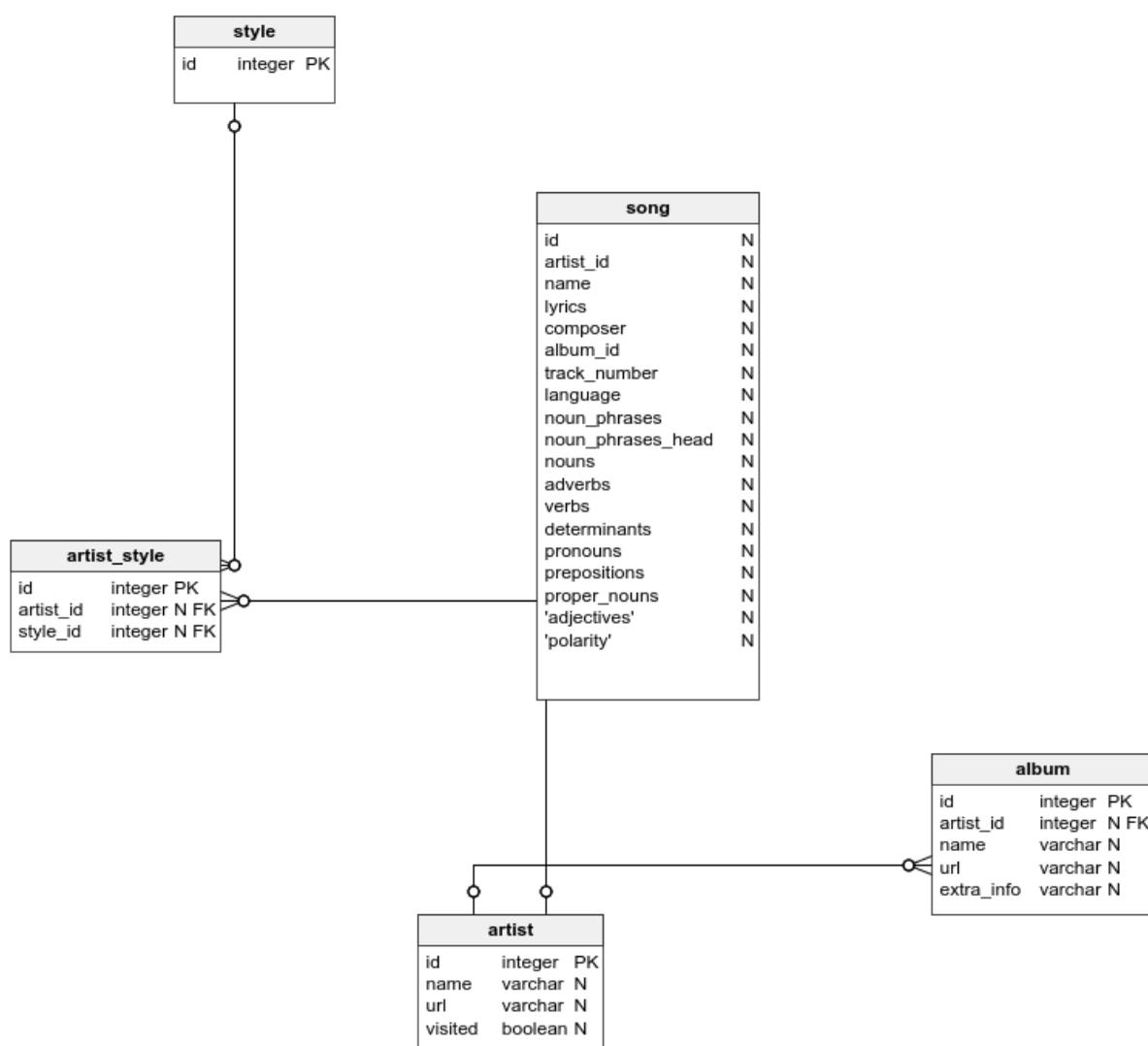
6 É permitida somente a visualização no site das letras de músicas encontradas aqui, vedada sua reprodução através de quaisquer outros meios (Lei 9610/98). Todas as letras de músicas são propriedade dos seus respectivos autores e divulgadas somente para fins educacionais

dos *crawlers*, programas de computador com o objetivo de realizar a captura contínua e completa dos dados da fonte escolhida. Os capturadores foram desenvolvidos utilizando-se a linguagem Python⁷ e os dados, armazenados em um banco de dados relacional Sqlite,⁸ escolhido pela simplicidade e portabilidade, o que será de extrema vali

a na disponibilização do banco de dados para a comunidade científica.

Após a captura e estruturação, enriquecemos as informações originais das letras com informações de marcação morfossintática para o português e o resultado da análise de sentimento. A figura 1 mostra o schema do banco MUSICAS.SQLITE.

Figura 1 - Banco de dados relacional MUSICAS.SQLITE



Fonte: Os autores, 2017

7 <https://www.python.org/>
 8 <https://www.sqlite.org/>

Os dados originais das letras de música, agregados às informações adicionadas pela marcação morfossintática e a análise de sentimento perfazem 3.2 GB, contando com 1616 letras de música em 38 línguas. O quadro 1 a seguir ilustra os campos das tabelas do banco de dados, e uma descrição completa se encontra nos anexos 1 e 2 deste documento.

Quadro 1 - Descrição dos campos da Tabela

Tabela	Campos
<i>album</i>	id, artist_id, name, url, extra-info
<i>artist</i>	id, name, url, visited
<i>artist_style</i>	id, artist_id, style_id
<i>song</i>	id, artist_id, name, lyrics, composer, album_id, track number, language, noun_phrases, noun_phrases_head, nouns, adverbs, verbs, determinants, pronouns, prepositions, proper_nouns, adjectives, polarity
<i>style</i>	Id, name

Fonte: Os autores, 2017

Percebe-se, pelo quadro 1, a presença não somente de campos relativos a identificação da letra, mas também um vasto conjunto de dados de natureza linguística (língua, sintagma nominal, substantivo, advérbio, verbos, determinantes, pronomes etc.) e um dado relativo a polaridade semântica (polaridade). A construção deste banco será detalhada a seguir.

Após finalizar a captura dos dados do sítio anteriormente aludido, procedeu-se ao enriquecimento dos campos derivados da análise morfossintática, que se limitou neste momento somente às letras em português do Brasil. Numa agenda futura, que dará continuidade a esta investigação, está prevista a análise morfossintática das letras em outras línguas, também presentes no banco. A tabela 1 a seguir ilustra as quantidades de músicas por língua, para as principais línguas presentes no banco de dados.

Tabela 1 - Quantidade de música por língua

Inglês	765696
Português	531222
Espanhol	96630
Francês	39915
Sueco	35477
Italiano	28205
Holandês	27102
Alemão	25559

Fonte: Os autores, 2017

A identificação da linguagem foi realizada utilizando-se o pacote *language detection*, disponível para a linguagem de programação Python, que possui uma taxa média de mais de 90% de acertos.

Esta etapa de caráter linguístico, realizada nas letras de música em português, obedeceu aos passos descritos a seguir. Primeiramente, procedeu-se a marcação morfossintática das letras de música em português usando o *parser* “Palavras” (BICK, 2000). Em seguida, extraíram-se os sintagmas nominais das letras utilizando a saída do *parser* e *scripts* adicionais na linguagem PERL. Estes foram incorporados às informações sintáticas ao banco de dados MUSICAS.SQLITE. Estando os dados linguísticos marcados no léxico, iniciou-se a etapa de classificação segundo a polaridade emocional, que envolveu a identificação de informações subjetivas veiculadas pelas letras de música utilizando técnicas de Análise de Sentimento. Para tanto, foram utilizados dois léxicos polarizados: o SENTILEX (SILVA; CARVALHO; SARMENTO, 2012), e o OPLEXICON (SOUZA; VIEIRA; Buseti.; CHISHMAN; ALVES, 2012); ambos contendo vocábulos e polaridades associadas (com escala Likert simplificada - positiva, neutra, negativa). O léxico SENTILEX foi desenvolvido a partir do português variante de Portugal e o OPLEXICON para o português na variante brasileira. A heurística de aplicação utilizada valeu-se da soma das polaridades das palavras presentes em pelo menos um destes vocabulários polarizados. Caso o vocábulo existisse nos dois léxicos, escolhia-se a informação do OPLEXICON, pela evidente preferência dos sentidos atribuídos no escopo do português do Brasil. A aplicação da técnica de Análise de Sentimento resultou na alimentação de forma automática do campo relativo a polaridade emocional (*polarity*) da tabela *Song*, como visto no quadro 1. Nesta pesquisa adotamos a abordagem lexical para a

aplicação de AS. Segundo Benvenuto, Ribeiro e Araujo, o método lexical utiliza

listas e dicionários de palavras associadas a sentimentos específicos. Apesar de não dependerem de dados rotulados para treinamento, a eficiência dos métodos léxicos está diretamente relacionada a generalização do vocabulário utilizado, para os diversos contextos existentes (BENVENUTO, RIBEIRO, ARAUJO, 2015, p. 5).

A pesquisa no banco de dados MUSICAS.SQLITE é realizada por meio de expressões em SQL. Por exemplo, para recuperar, em ordem decrescente de polaridade emocional, as letras do compositor Cazuzza e os dados associados às letras, poderíamos ter como exemplo a expressão de busca descrita no quadro 2.

Quadro 2 - Expressão de busca em SQL

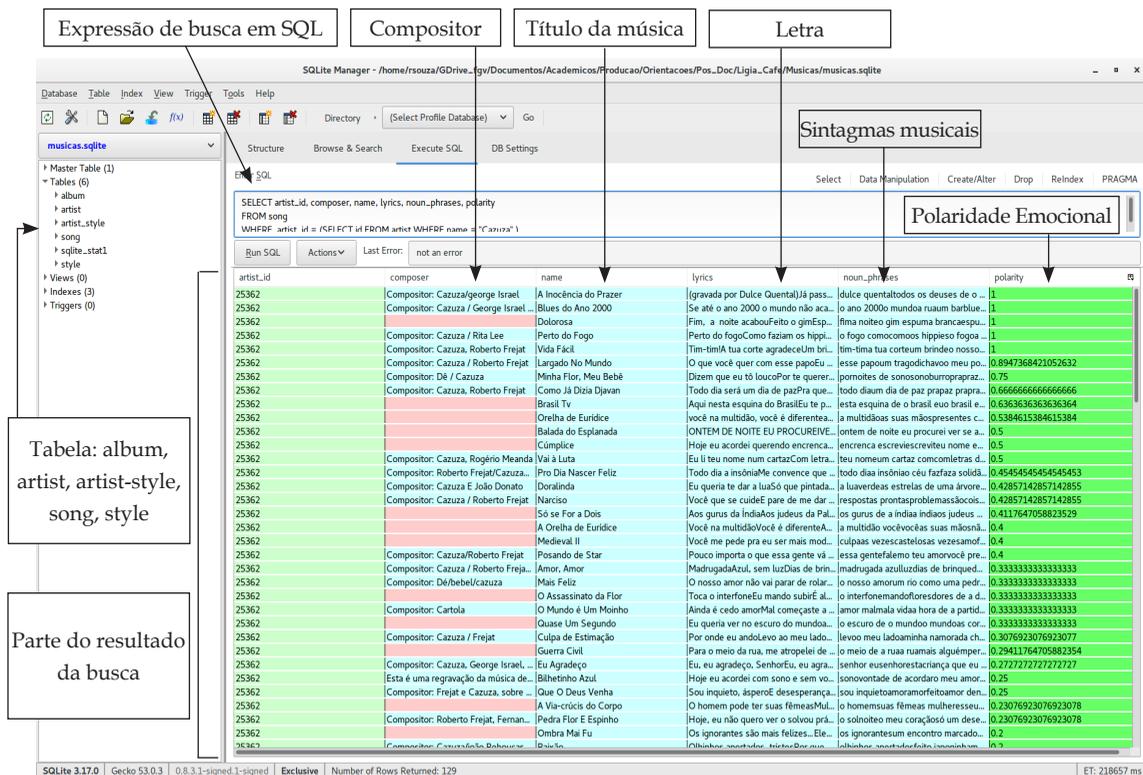
```
SELECT artist_id, composer, name, lyrics, noun_phrases, polarity
FROM song
WHERE artist_id = (SELECT id FROM artist WHERE name = "Cazuzza")
ORDER BY polarity
DESC;
```

Fonte: Os autores, 2017

O resultado é ilustrado na figura 2. No centro da figura 2, encontra-se o espaço reservado para pesquisa no banco MUSICAS.

SQLITE. Na última coluna do lado direito pode ser visualizada a polaridade emocional em ordem decrescente das músicas do compositor Cazuzza.

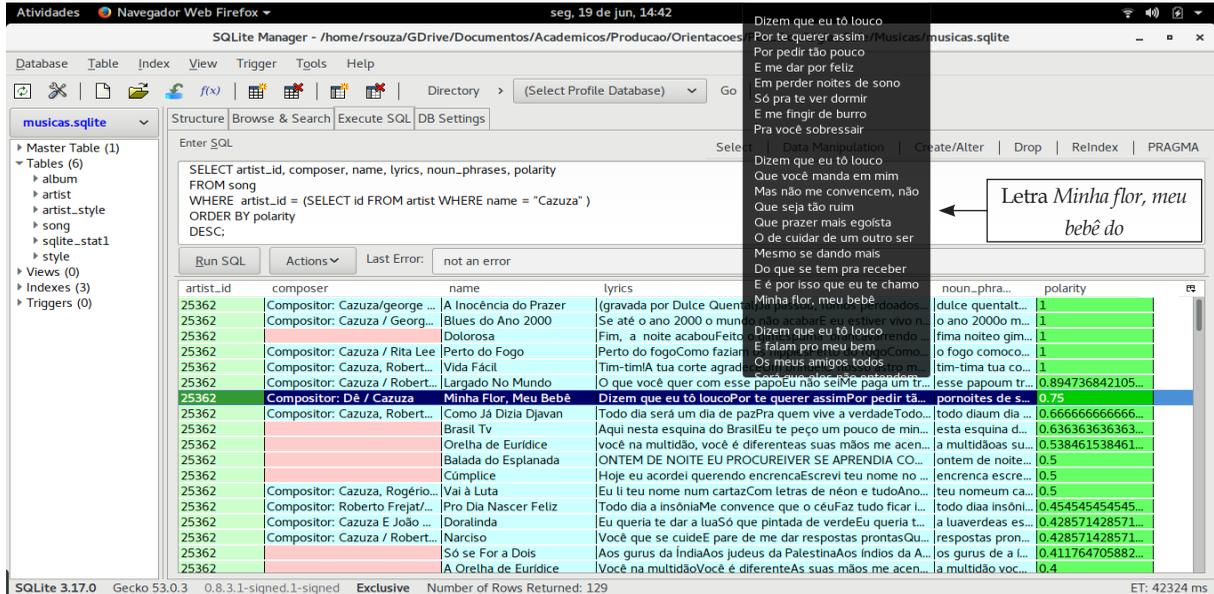
Figura 2: Exemplo de busca relacionada a letra do compositor Cazuzza



Fonte: Os autores, 2017

Passando o mouse sobre as colunas *lyrics* (letras) ou *noun_phrases* (sintagma nominal) é possível ampliar a informação. No caso da figura 3, ampliamos a letra *Minha flor, meu bebê*.

Figura 3: Exemplo de busca relacionada a letra *Minha flor, meu bebê* do compositor Cazuza



Fonte: Os autores, 2017

Uma simples análise do ranking permite verificar que, embora se façam todas as ressalvas em relação à empiricidade e grande dose de subjetivismo do método de análise de sentimentos, os termos utilizados nas canções e o sentimento que estas carregam estão diretamente vinculados. Acreditamos que o vocabulário escolhido pelo compositor é o veículo daquilo que deseja exprimir afetivamente. Conseqüentemente, a direção semântica (ou polaridade emocional) que o compositor almeja alcançar influencia e é influenciada pelas escolhas que ele faz no vasto vocabulário da língua. No exemplo acima, as palavras de “Minha Flor, meu bebê” é evidentemente mais “positiva” que a letra de “Pedra, Flor e Espinho”, que tem palavras como “Obscuro”, “Medo” e “Espinho”, consideradas de polaridade mais negativa. Deve-se ressaltar, porém, novamente, que músicas, como expressões artísticas, são subjetivas e deve-se evitar a afirmação de que uma letra tem ou não sentimento negativo. Nosso banco de dados apenas aponta para indícios baseados nas palavras.

3 CONSIDERAÇÕES FINAIS

Esta investigação pretende contribuir para o avanço de pesquisas sobre recuperação automática da informação musical e oferece como produto um banco de dados robusto à disposição da comunidade científica. Este banco pode ser acessado no link indicado

Apesar de oferecermos alguns poucos exemplos, a riqueza deste campo que se abre para a organização de informação advirá do uso deste recurso por pesquisadores de nossa área, dos quais esperamos que corroborem as possibilidades de constatação destes fenômenos linguísticos, de mensuração de aspectos afetivos através do processamento automático da linguagem, especificamente via a técnica de Análise de Sentimento. Os autores também se colocam abertos para escutar propostas de pesquisas com o recurso ora oferecido.

SENTIMENT ANALYSIS APPLIED TO THE STUDY OF MUSIC LYRICS

ABSTRACT *Music as an information object is a complex topic in the field of knowledge organization (KO), especially when the modeling and analysis occurs in an automated way. One of the techniques, derived from the field of Artificial Intelligence and Natural Language Processing (NLP) that has emerged as promising in this arena is the Sentiment Analysis (SA), or, alternatively, Opinion Mining; used to identify opinions and emotions in texts, evaluating them as positive or negative. This research present an experience of sentiment analysis with music lyrics. After building a database of lyrics in several languages made from crawling lyrics websites, the Brazilian Portuguese lyrics were morphosyntactically analysed. Then, the SA technique was applied to classify the corpus according to the emotional dimension, which resulted in a database of 3.2 GB, with 1616 lyrics in 38 languages, enriched with Part of Speech (POS) tagging and sentiment polarity for Portuguese. This experience has shown that the affective aspect can be measured or at least indicated by natural language processing, specifically via SA techniques. The results lead to the conclusion of the effectiveness of SA for music lyrics, attesting its applicability and contribution to the advancement of research in the field to the KO in this domain*

Keywords: *Music Lyrics. Sentiment Analysis. Natural Language Processing.*

REFERÊNCIAS

BALAGE FILHO, Pedro P.; PARDO, Thiago A. S.; ALUISIO, Sandra M. An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In: BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 9, 2013, Fortaleza, **Anais...**, 2013.

BENVENUTO, Fabrício; RIBEIRO, Fabrício; ARAÚJO, Matheus. Métodos para análise de sentimentos em mídias sociais. In: **Short course in the Brazilian Symposium on Multimedia and the Web (Webmedia)**. Manaus, 2015. Disponível em: < <http://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf>>. Acesso em: 18 fev. 2017.

BICK, Eckhard. **The parsing system Palavras: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**. Aarhus, Dinamarca: Aarhus University Press, 2000. 505 p. Disponível em: < <http://beta.visl.sdu.dk/%7EEckhard/pdf/PLP20-amilo.ps.pdf>> Acesso em: 20 fev. 2017.

CAFÉ, Lígia; SOUZA, Renato Rocha. Sentiment Analysis and Knowledge Organization:

an overview of the international literature. **Knowledge Organization Journal**. Special Issue on New Trends for Knowledge Organization, v. 44, n.3, p. 199-214, 2017.

HATZIVASSILOGLOU, Vasileios; KATHLEEN R. McKeown. Predicting the semantic orientation of adjectives. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 35, 1997, p. 174-181, **Anais eletrônicos...**, 1997. Disponível em: < http://delivery.acm.org/10.1145/980000/979640/p174-hatzivassiloglou.df?ip=200.193.93.91&id=979640&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=671530387&CFTOKEN=98242318&acm__=1430772585_20_9e094b2668aa2693d1b068790f54b2> Acesso em: 18 fev. 2017.

JAIN, Tejashri Inadarchand; NEMADE, Dipak. Recognizing contextual polarity in phrase-level sentiment analysis. **International Journal of Computer Applications**, v. 7, n.5, p.12-21, 2010.

OUDENNE, Ashley M.; CHASINS, Sarah E. **Identifying the Emotional Polarity of Song Lyrics through Natural Language Processing**. Swarthmore, EUA, s.d. 13p. Relatório final

apresentado no CPSC 65: Natural Language Processing, Swarthmore College. Disponível em: <http://www.sccs.swarthmore.edu/users/11/aoudenn1/Ashley_M.Oudenne/Research_files/emotionalPolarity.pdf> Acesso em: 18 fev. 2017.

PANG, Bo; LEE, Lillian. Opinion mining and sentiment analysis. **Foundation and trends in information retrieval**. v. 2, n. 1-2, p.1-135, 2008.

RIGO, Sandro J. et al. Abordagem linguística para identificação da dimensão afetiva expressa em textos de Ambientes Virtuais de Aprendizagem: um léxico da emoção. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2, 2013, Campinas, **Anais...**, 2013.

SILVA, M.; CARVALHO, C.; SARMENTO, L. Building a Sentiment Lexicon for Social Judgement Mining. In: CASELI, h. et al. (Orgs). COMPUTATIONAL PROCESSING OF THE PORTUGUESE LANGUAGE, INTERNATIONAL CONFERENCE, PROPOR, 10, 2012, Coimbra,

Portugal, April 17-20, 2012. Berlin: Springer Berlin Heidelberg, 2012. p. 218-228. (Lecture Notes in Computer Science, 7243).

SOUZA, M.; VIEIRA, R.; BUSETTI, D.; CHISHMAN, R.; ALVES, I. M. Construction of a Portuguese Opinion Lexicon from multiple resources. 8th Brazilian Symposium in Information and Human Language Technology, 2012.

WILSON, Theresa; WIEBE, Janyce; HOFFMANN, Paul. Recognizing contextual polarity in phrase-level sentiment analysis. In: **Proceedings of the conference on human language technology and empirical methods in natural language processing**. Association for Computational Linguistics, 2005. p. 347-354.

ANEXO 1

Documentação do Banco de Dados Musicas.sqlite

1. Tabela album

Column name	Type	Properties	Description
id	integer		
artist_id	integer	null	
name	varchar	null	
url	varchar	null	
extra_info	varchar	null	

2. Tabela artist

Column name	Type	Properties	Description
id	integer		
name	varchar	null	
url	varchar	null	
visited	boolean	null	

3. Tabela artist_style

Column name	Type	Properties	Description
id	integer		
artist_id	integer	null	
style_id	integer	null	

4. Tabela song

Column name	Type	Properties	Description
id		null	
artist_id		null	

name		null	
lyrics		null	
composer		null	
album_id		null	
track_number		null	
language		null	
noun_phrases		null	
noun_phrases_head		null	
nouns		null	
adverbs		null	
verbs		null	
determinants		null	
pronouns		null	
prepositions		null	
proper_nouns		null	
'adjectives'		null	
'polarity'		null	

5. Tabela style

Column name	Type	Properties	Description
id	integer		
name	varchar	null	

ANEXO 2

Script para geração do Banco musicas.sqlite

```
-- tables
-- Table: album
CREATE TABLE album (
  id integer NOT NULL CONSTRAINT album_pk PRIMARY KEY,
  artist_id integer,
  name varchar,
  url varchar,
  extra_info varchar,
  CONSTRAINT AK_0 UNIQUE (url),
  CONSTRAINT FK_3 FOREIGN KEY (artist_id)
  REFERENCES artist (id)
);

-- Table: artist
CREATE TABLE artist (
  id integer NOT NULL CONSTRAINT artist_pk PRIMARY KEY,
  name varchar,
  url varchar,
  visited boolean,
  CONSTRAINT AK_1 UNIQUE (url),
  CONSTRAINT CHECK_0 CHECK (( visited IN ( 0 , 1 )))
);

-- Table: artist_style
CREATE TABLE artist_style (
  id integer NOT NULL CONSTRAINT artist_style_pk PRIMARY KEY,
  artist_id integer,
  style_id integer,
  CONSTRAINT AK_2 UNIQUE (artist_id, style_id),
  CONSTRAINT FK_1 FOREIGN KEY (artist_id)
  REFERENCES artist (id),
  CONSTRAINT FK_5 FOREIGN KEY (style_id)
  REFERENCES style (id)
);

-- Table: song
CREATE TABLE song (
  id,
  artist_id,
  name,
  lyrics,
  composer,
  album_id,
  track_number,
  language,
  noun_phrases,
  noun_phrases_head,
  nouns,
  adverbs,
  verbs,
  determinants,
  pronouns,
  prepositions,
  proper_nouns,
  'adjectives',
  'polarity'
);

-- Table: style
CREATE TABLE style (
  id integer NOT NULL CONSTRAINT style_pk PRIMARY KEY,
  name varchar
);

-- End of file.
```