

TRABALHO FINAL WEB MINING

Aluna: Isadora Berg

29/03/2016

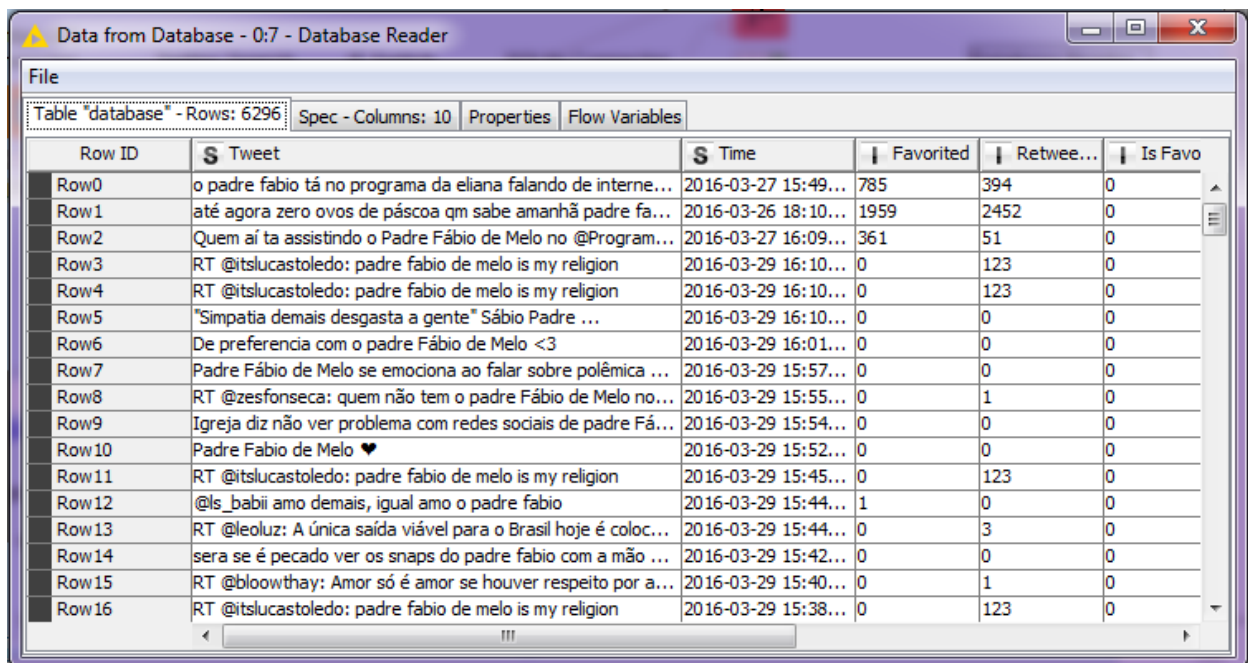
1. Contextualização da proposta do trabalho

Este trabalho propõe coletar tweets sobre o Padre Fábio de Melo. Sabe-se que o Padre Fábio, antes um ícone da igreja católica, hoje têm entretido muitos seguidores do snapchat e twitter com vídeos e pensamentos humorísticos. Desta forma, seria interessante analisar quais os termos estão mais relacionados a esta personalidade, bem como descobrir como as pessoas tem se sentido em relação a este lado humorístico do Padre. Ainda vamos destacar qual o usuário do twitter tem tido maior influência no que diz respeito às citações do Padre Fábio.

2. Descrição do workflow no KNIME

Para fazer as análises propostas, foi utilizado o programa KNIME. Abaixo descreveremos o workflow construído, que se encontra completo em outro arquivo anexado neste trabalho.

Na primeira parte do trabalho consta a leitura do API Twitter. Consideramos 100mil tweets com o termo “Padre Fabio”, porém o programa retornou apenas 6.296 tweets. O node “Data Base Reader” mostra a base de dados construída com esses tweets.



Row ID	Tweet	Time	Favorited	Retwee...	Is Favo
Row0	o padre fabio tá no programa da eliana falando de interne...	2016-03-27 15:49...	785	394	0
Row1	até agora zero ovos de páscoa qm sabe amanhã padre fa...	2016-03-26 18:10...	1959	2452	0
Row2	Quem aí ta assistindo o Padre Fábio de Melo no @Program...	2016-03-27 16:09...	361	51	0
Row3	RT @itslucastoledo: padre fabio de melo is my religion	2016-03-29 16:10...	0	123	0
Row4	RT @itslucastoledo: padre fabio de melo is my religion	2016-03-29 16:10...	0	123	0
Row5	"Simpatia demais desgasta a gente" Sábio Padre ...	2016-03-29 16:10...	0	0	0
Row6	De preferencia com o padre Fábio de Melo <3	2016-03-29 16:01...	0	0	0
Row7	Padre Fábio de Melo se emociona ao falar sobre polêmica ...	2016-03-29 15:57...	0	0	0
Row8	RT @zesfonseca: quem não tem o padre Fábio de Melo no...	2016-03-29 15:55...	0	1	0
Row9	Igreja diz não ver problema com redes sociais de padre Fá...	2016-03-29 15:54...	0	0	0
Row10	Padre Fabio de Melo ♥	2016-03-29 15:52...	0	0	0
Row11	RT @itslucastoledo: padre fabio de melo is my religion	2016-03-29 15:45...	0	123	0
Row12	@ls_babii amo demais, igual amo o padre fabio	2016-03-29 15:44...	1	0	0
Row13	RT @leoluz: A única saída viável para o Brasil hoje é coloc...	2016-03-29 15:44...	0	3	0
Row14	sera se é pecado ver os snaps do padre fabio com a mão ...	2016-03-29 15:42...	0	0	0
Row15	RT @bloowthay: Amor só é amor se houver respeito por a...	2016-03-29 15:40...	0	1	0
Row16	RT @itslucastoledo: padre fabio de melo is my religion	2016-03-29 15:38...	0	123	0

A segunda parte do trabalho consiste no enriquecimento dos dados. Como muitos nodes de processamento de textos estão disponíveis apenas para a língua inglesa, o trabalho ficou mais restrito nesta parte, porém tentou-se não comprometer os resultados.

Desta forma, utilizou-se primeiramente o filtro *regex* para exclusão de nomes de usuários, que apareciam no documento devido aos retweets. Em seguida foram excluídas as pontuações e termos não alfa-numéricos. Por último utilizou-se o node *stop word* que exclui palavras sem sentido literário. O node foi utilizado duas vezes, sendo que na primeira ele considera uma lista personalizada de palavras a excluir e na segunda ele utiliza a lista pronta do programa.

Após esta etapa, fez-se necessário substituir algumas strings de mesmo significado para uma única string. Por exemplo, as palavras “fábio”, “FABIO”, “fabio”, foram todas substituídas por “Fábio”. Isto para melhorar o desempenho da agregação de termos e identificação da relevância destes.

Na terceira parte construiu-se a tag cloud, que mostra os termos citados mais relevantes. A relevância aqui foi considerada segundo o IDF.



A parte seguinte do workflow contém a análise de sentimento. Neste sentido, procurou-se identificar quais os adjetivos e substantivos têm sido correlacionados com o nome Padre Fábio, e se esses termos têm sido usados de forma positiva ou negativa. O resultado se encontra no tag cloud

abaixo, em que as palavras em verde têm o significado positivo, e as palavras em vermelho têm significado negativo.

definitivamente amado imprensa crise esquecer preciso
odeio surpreender escuro paz gosto luz lindo chama amo
odiosa intimidade tipo pro respeito amante inferno
bonito crush loucura simpatia
vontade viver precisa legal
amizade vivo travesti
gloss problema tenso sorrindo
sabedoria errado melhor pecado perda
infelizmente universal pena culpa saudade chorar superar

Na última parte do workflow, procuramos o usuário mais retwittado sobre o assunto Padre Fábio e ainda qual o tweet que gerou mais replicações. Os resultados se encontram na tela abaixo.

Usuário: @governordie

Quantidade de retweets: 10.465

Row ID	Tweet	Time	Favorited	Retweeted
Row 4638	RT @governordie: o cara é bonitrico bem humorado é zueiro no twitter cara que você quer pra namorar o crush só tem um problema é o Pa...	2016-03-22 19:45...	0	10465

Tweet na fonte:



mulan das bad
@governordie



Follow

o cara é bonito
rico
bem humorado
é zueiro no twitter
o cara que você quer pra namorar
é o crush
só tem um problema
é o Padre Fábio de Melo

View translation

RETWEETS
10,465

LIKES
6,870



3. Conclusões

Com este trabalho pudemos identificar quais termos estão sendo mais correlacionados com o Padre Fábio, e qual o sentimento as pessoas têm em relação a esta personalidade. O tag cloud nos permitiu visualizar palavras como “snaps”, “religion”, “casamento”. Buscando estas palavras entre os tweets coletados, descobrimos que as pessoas consideram o usuário do Padre Fábio o melhor usuário do snapchat. Consideram que o Padre é a sua religião, em tweets como “Padre Fabio is my religion”. E há ainda quem sonha casar com o Padre.

A análise de sentimento resultou em palavras negativas, destacas em vermelho. Palavras como “pecado”, “loucura” e “crush”. Porém ao pesquisarmos o contexto em que essas palavras ocorrem, descobrimos que seu sentido não é negativo. Por exemplo, a palavra “pecado” aparece em tweets declarando que o usuário quer cometer “pecado” com o padre. Ainda, que sentem um “crush” pelo Padre. E que este vai a loucura no snap, com posts humorísticos.

Além disso, pudemos também identificar o usuário que teve o post mais retwittado, e confirmamos a impressão positiva que os usuários do twitter têm sobre o Padre, uma vez que o conteúdo do tweet elogia o Padre.

Como proposta para complemento deste trabalho, poderia-se minerar melhor os textos dos tweets, excluindo ainda mais palavras sem sentidos literários. A análise de sentimento poderia ser enriquecida, se o dicionário de palavras contivesse palavras acentuadas. Ainda nesse sentido, uma melhor classificação das palavras poderia ser feita, a exemplo da palavra “simpatia” que foi pré-classificada como negativa, quando seu sentido é positivo. Poderia-se ainda criar uma rede contendo o próprio usuário do Padre Fábio e demais usuários, analisando estas relações.