1

**Projeto Interdisciplinar BD-ITAC**

**CE-240 - Projeto de Sistemas de Banco de Dados**

Prof. Dr. Adilson Marques da **Cunha**

**Requisitos TS#01**

**Colaboração inicial: Henrique Duarte Borges Louro**

Requisitos:

1. O Sistema poderá monitorar uma ou mais crises ao mesmo tempo;
2. Cada crise cadastrada será uma instância ativa, ou não, do sistema (thread), executadas em processamento paralelo;
3. O Sistema contará com a possibilidade de monitorar várias APIs de coleta de dados ao mesmo tempo e por crise;
4. Cada crise terá suas APIs selecionadas, ativadas ou não;
5. As APIs ativadas monitorarão as mensagens em tempo real entregando seu conteúdo para as instâncias ativas de crise que estiverem associadas;
6. Cada instância de crise verificará, num primeiro momento, se a mensagem recebida da API associada é relevante ou não, de acordo com os critérios previamente cadastrados, descartando-a caso não seja;
7. A verificação de cada mensagem recebida se dará em duas etapas: Identificação, que verificará se a mensagem tem a ver com a crise ou não e posteriormente a Classificação, que indicará, entre outras possibilidades, o momento em que a crise se encontra, os sentimentos a ela atribuídos e possíveis ações e/ou comunicações necessárias sobre o evento;
8. Essas verificações (Identificação e Classificação) também deverão ser feitas em instâncias (threads) permitindo o processamento em paralelo. Justifica-se tal procedimento pois, os esforços computacionais para uma e outra ação poderão ser diferentes em cada momento e ações poderão já estarem sendo executadas, logo após a identificação da crise pela mensagem, enquanto ela aguarda o processamento da sua classificação, que por sua vez, deverá consumir mais tempo de processamento;
9. Ao se identificar que uma mensagem é relevante para a crise em questão, será gravada numa base de dados e estará disponível para a sequência de classificação e outras ações do sistema;
10. Os critérios associados a cada crise serão feitos através de palavras chaves, no caso de APIs de Redes Sociais, que funcionarão através de mensagens recebidas, monitoramento de sensores no caso de IoT e informações públicas de monitoramento recebidas de órgãos públicos;
11. No caso de critérios das APIs de redes sociais, as palavras associadas serão separadas em três tipos: substantivos, verbos e adjetivos;
12. Os substantivos e verbos identificados nas mensagens, estarão associados ao tipo de evento e servirão para selecionar a mensagem como relevante ou não para a crise em questão;
13. Os verbos, além da função anterior, poderão identificar se o evento ocorreu, está ocorrendo ou ocorrerá, o que permitirá ações de prevenção e/ou contingência, assim que a crise for identificada e comunicada;
14. Os adjetivos identificados poderão informar qual o grau de gravidade de um evento (sentimento) e até mesmo pautar as ações necessárias de intervenção; (verificar onde se encaixam os aumentativos e diminutivos???)
15. Ao se cadastrar um critério para as APIs Sociais (palavra chave) o sistema deverá, através de metadados, no caso de substantivo, associar seu plural e possíveis sinônimos e no caso de verbos, as conjugações verbais possíveis, necessárias para identificar o tempo de ocorrência do evento. Tal ação do sistema, visa minimizar a necessidade de intervenção humana para cadastrar os critérios;
16. É importante que os critérios esteja todos cadastrados previamente e que não haja necessidade de gera-los em tempo de processamento, pois isso diminuirá o tempo de resposta do sistema na hora da identificação e classificação da mensagem recebida;
17. Tais critérios serão carregados para a memória ao se ativar uma crise e suas respectivas APIs e ali estarão para as possíveis comparações, o que tornará a velocidade de processamento muito maior, uma vez que não haverá necessidade de acesso ao BD a cada verificação que se faça nas palavras das mensagens;
18. Com relação às mensagens, serão quebradas em palavras através dos espaços nelas contidos. Uma outra verificação que se faz necessário é a pontuação existente, que também deverá ser retirada. Um caso especifico que deverá ser verificado é quando uma pontuação não se encontrar no final de uma palavra e sim no meio, o que indicará que se trata de duas palavras às quais o usuário não utilizou espaço para separa-las;
19. Palavras menores que três letras deverão ser descartadas;

Questões:

Ficam aqui outras questões para serem pensadas e discutidas quanto à sua relevância e possibilidade de implementação ou não no sistema:

1. É necessário um estudo com relação a erros de grafia. Como serão tratados? Existem algoritmos prontos para isso, que dependem de uma base extensa de possíveis palavras. Será que o esforço computacional envolvido compensa o trabalho e não inviabilizará os tempos necessários para identificação e classificação das mensagens?
2. Como tratar as abreviações encontradas nas mensagens? Teremos um dicionário de termos? Mais uma vez a questão: vale a pena o esforço computacional envolvido?
3. No caso das hashtags poderemos ter várias palavras unidas para forma-la. Para tratar isso, o sistema precisaria varrer todo um banco de dados, com o máximo de palavras existentes na língua utilizada, fazendo buscas por todas as associações possíveis de três ou mais letras, da esquerda para a direita, para identificar cada uma das palavras utilizadas na composição da hashtag. Será que compensará o esforço computacional envolvido?
4. Como serão tratadas as informações recebidas de Iot? Para cada tipo de sensor serão necessários critérios específicos?
5. Como serão tratadas as informações recebidas de Órgãos Públicos? Cada interação com esses sistemas necessitará de parâmetros específicos?