CONSTRUÇÃO DE UMA PLATAFORMA ONLINE PARA MINERAÇÃO DE OPINIÃO EM PORTUGUÊS

Sérgio Ricardo G. B. Filho, Thiago A. S. Pardo

Núcleo Interinstitucional de Linguística Computacional (NILC), Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo

sergiobarbosa@usp.br, taspardo@icmc.usp.br

Objetivos

No presente projeto pretendemos desenvolver uma API que reunisse as ferramentas de análise de sentimentos produzidas pelo projeto OPINANDO e que permitisse o funcionamento integrado delas para processar revisões de produtos da web e produzir resumos. Pretendemos também desenvolver uma plataforma online para a utilização dessas funcionalidades.

Métodos e Procedimentos

Implementamos um *crawler* de opiniões para conseguir extrair revisões da *web* a partir do nome do produto desejado. Iniciamos então a adaptação das ferramentas do projeto OPINANDO para torná-las compatíveis com o funcionamento da API e, em seguida, fizemos a integração de forma a tornar possível utilizá-las em conjunto.
Dessa maneira, para gerar o resumo das revisões extraídas, elas são processadas através de um pipeline de sentimento (Figura 1) onde cada etapa corresponde ao uso de uma ferramenta diferente.

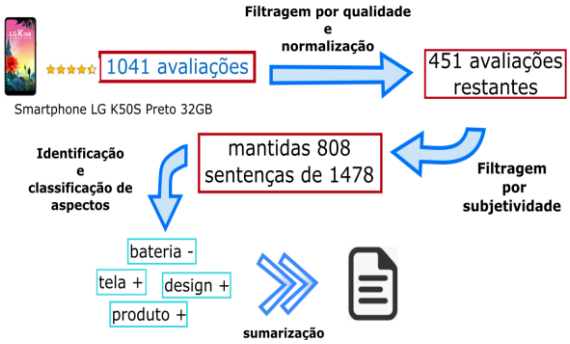

Figura 1: etapas do pipeline de sentimento

Resultados

No total foram inseridas, até então, 6 ferramentas na API. Dentre elas estão o normalizador Enelvo [1] e o sumarizador Opizer [2]. A API criada foi disponibilizada para download através do github. Em seu formato atual, ela facilita a inserção de novas ferramentas, necessitando apenas que suas entradas e saídas sejam feitas utilizando arquivos em formato *json*.

Desenvolvemos a plataforma *web* intitulada OpinionSpot (Figura 2) que utiliza a API com base. Ela pode ser acessada através do link http://nilc.icmc.usp.br/opinionspot/.

Figura 2: recortes de várias visões da interface *web*

Encontramos dificuldade em manter a disponibilidade do *crawler* na versão *web*, necessitando ainda de um suporte constante para manter funcionando e de melhorias para encontrar um maior número de revisões sem grande impacto no tempo de resposta da aplicação.

Conclusões

As aplicações desenvolvidas facilitam muito o acesso e utilização das ferramentas do OPINANDO ao permitir reunir todas em um só lugar. A interface *web* as torna ainda disponíveis para fácil utilização do público em geral.

Referências Bibliográficas

[1] Bertaglia, Thales Felipe Costa, and Maria das Graças Volpe Nunes. Exploring Word Embeddings for Unsupervised Textual User-Generated Content Normalization. In the Proceedings of the 2nd Workshop on Noisy User-generated Text, pp. 112-120. 2016.
[2] López Condori, R.E. Sumarização automática de opiniões baseada em aspectos. Dissertação de Mestrado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos-SP, July, 160p. 2015.