

SOCIAL MEDIA TEXT PROCESSING AND SEMANTIC ANALYSIS FOR SMART CITIES

João Pereira

Project/Dissertation performed under the supervision of *Prof. Rosaldo Rossetti*
in *LIACC*

1. Motivation

With the rise of Social Media, people obtain and share information almost instantly on a 24/7 basis. Many research areas have tried to gain valuable insights into these large volumes of freely available user generated content. The research areas of intelligent transportation systems and smart cities are no exception. However, extracting meaningful and actionable knowledge from user generated content is a complex endeavor. First, each social media service has its own data collection specificities and constraints, second the volume of messages/posts produced can be overwhelming for automatic processing and mining, and last but not the least, social media texts are usually short, informal, with a lot of abbreviations, jargon, slang and idioms.

2. Goals

In this dissertation, we try to tackle some of the aforementioned challenges with the goal of extracting knowledge from social media streams that might be useful in the context of intelligent transportation systems and smart cities. We designed and developed a framework for collection, processing and mining of geo-located Tweets. More specifically, it provides functionalities for parallel collection of geo-located

tweets from multiple pre-defined bounding boxes (cities or regions), including filtering of non complying tweets, text pre-processing for Portuguese and English language, topic modeling, and transportation-specific text classifiers, as well as, aggregation and data visualization.

3. Descrição do Trabalho

3.1. Línguas Obrigatórias

3.2. Cabeçalho de Identificação

3.3. Extensão do Artigo e Tipo de Letra

3.4. Dimensões das Margens

3.5. Corpo do Texto

3.6. Numeração das Secções

3.7. Figuras, Equações e Quadros

3.8. Numeração de Páginas

3.9. Codificação dos Caracteres

4. Conclusões

Referências