

On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing

Rich Elton Carvalho Ramalho, Anderson Almeida Firmino, Cláudio de Souza Baptista, Ana Gabrielle Ramos Falcão, Maxwell Guimarães de Oliveira, Fabio Gomes de Andrade

Cláudio de Souza Baptista, PhD - Federal University of Campina Grande



baptista@computacao.ufcg.edu.br

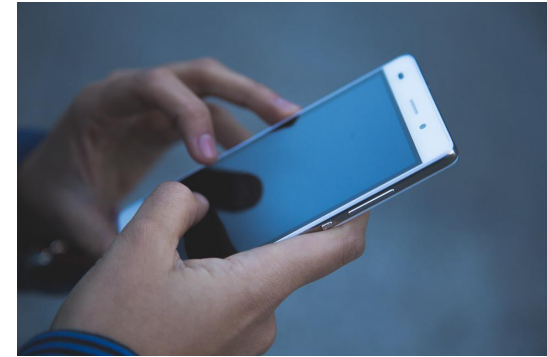
Cláudio de Souza Baptista is a Full Professor of the Computer Science Department of the Federal University of Campina Grande (UFCG), where he presides the Information Systems Laboratory (LSI), working with the Doctorate, Masters and Bachelor Degrees of Computer Science. He is a CNPQ's DTI-2A researcher, having published over 100 articles on national and international conferences and journals. His areas of interest include: multimedia, databases, geoprocessing and natural language processing.



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



- Citizens as sensors enable the engagement of society through technology to complain on urban issues;
- Although there are specialized geosocial networks for such, users' engagement decreases over time;
- Many relevant issues may not be identified or published, which reduces the effectiveness of these networks.



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



- Aiming to overcome this limitation, we propose an approach in which urban issues are automatically detected from a TV news program;



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



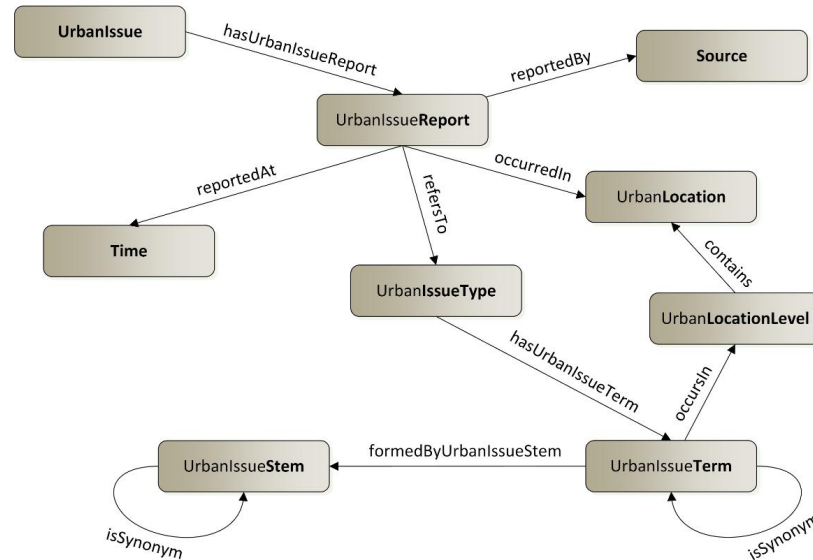
- Automated method for Extracting and structuring urban issues reported in TV News:
 - Web scraping method to extract the audio from video news;
 - Convert the audio into text using a speech recognition tool;
 - Use a gazetteer to perform geoparsing on the mentioned addresses and locations obtained from the Named Entity Recognition (NER) process, without preprocessing;
 - Custom preprocessing step comprising word capitalization, stopwords removal and lemmatization;
 - Use NER to obtain the named entities from the text;
 - Perform topic modeling to obtain the class of urban issues related to the text;
 - Finally, the urban issues are located into the Crowd4City geosocial network.

On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



- The UIDO Ontology:

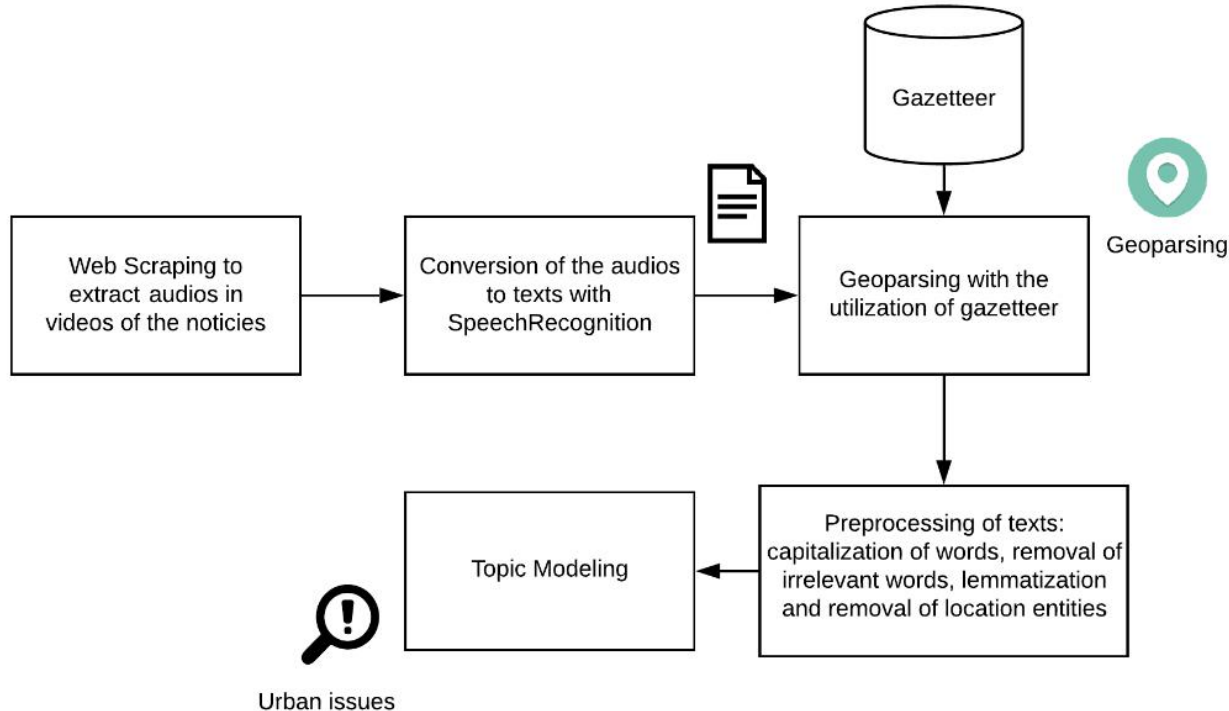
- Urban Issues Domain Ontology, developed using the Web Ontology Language (OWL) and the Protégé 5.0 system;



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



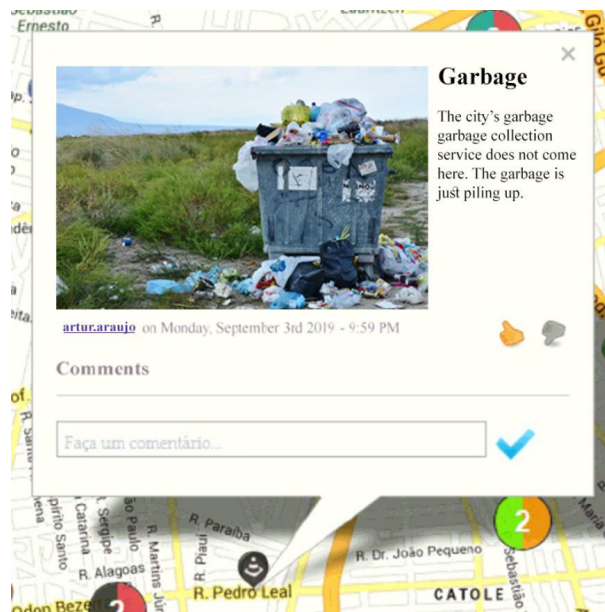
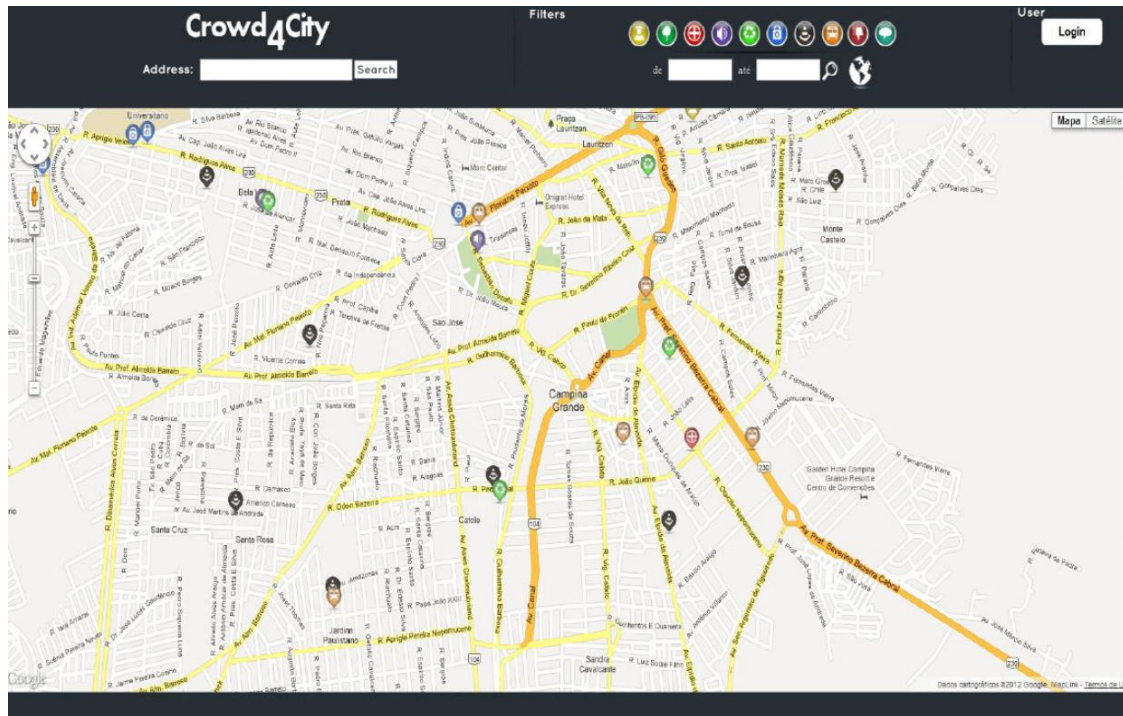
- Geoparsing



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



- Crowd4City



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



- Results - Topic Modeling

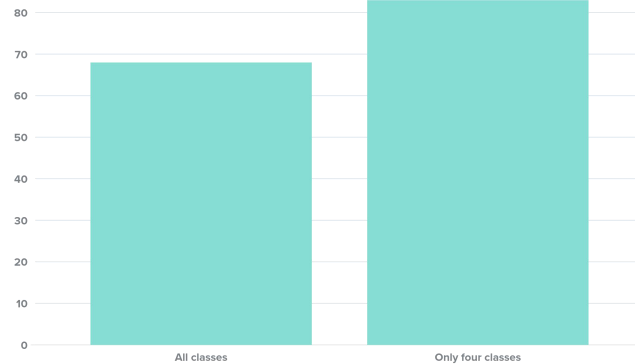
Class	Number of instances
Streets	308
Other (construction works)	236
Other (sanitation)	108
Education	91
Garbage	79
Other (traffic)	64
Complaints	26
Health	24
Security	17
Other (water shortage)	17
Public transportation	14
Forestation	5
Other	1

On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing

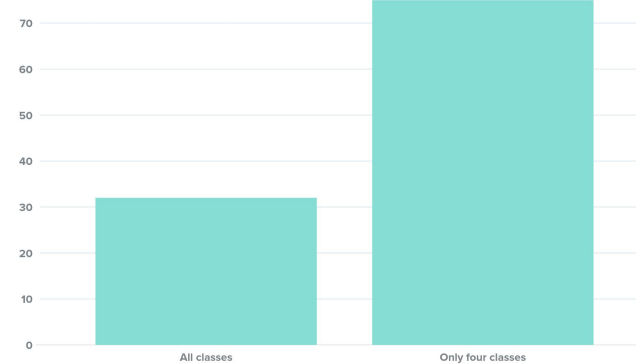


- Results - Topic Classification

XGBoost



Bi-LSTM



On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing



- Our approach is feasible and we manage to classify urban issues into four topics: mobility, sanitation, buildings and others.
- As future work, we plan to perform a performance analysis of geoparsing and a comparative study between topic modeling and supervised machine learning.

On classifying Urban Issues from TV News using Natural Language Processing and Geoprocessing

Rich Elton Carvalho Ramalho, Anderson Almeida Firmino, Cláudio de Souza Baptista, Ana Gabrielle Ramos Falcão, Maxwell Guimarães de Oliveira, Fabio Gomes de Andrade

Cláudio de Souza Baptista, PhD - Federal University of Campina Grande



baptista@computacao.ufcg.edu.br