

Implementation of a news event database

The project of implementation of a news event database is a data gathering and analysis system (DGAS), which is dedicated to tracking the European Media Monitor news aggregator, collecting the data, and performing multiple NLP-analysis operations. The system permanently scans the RSS feeds for news episodes, which are automatically formed by EMM and are based on various news sources. Each episode consists of numerous events, provided by EMM news partners.

DGAS fetches news events, including title, summary, publication date and other event's metadata. In addition, the application tries heuristically to retrieve the full text of the news for the further processing concerning various academic research. Besides the news' full text, the system is also looking for high-resolution images, which illustrate the news item.

News items, for them DGAS has succeeded to fetch the full text, undergoes NLP-analysis, which retrieves the keywords, named entities, such as locations, organisations, and persons, and performs the sentiment analysis to compute the overall news item polarity rank. All locations are provided with its geo points.

The final step of the news episode lifecycle is summarising its metadata. At the end of this stage, the system holds the metadata of the entire news episode, including the number of events, episode life expectancy, the lists of most frequent keywords, locations, organisations, and persons. To conclude, the main goal of DGAS is to build a system for data gathering and processing of news items, which can be used for the future research in different domains, such as data visualisation, an information retrieving, and data science.

Architecture, design, and implementation:

Michael Budnyatsky // <https://linkedin.com/in/budnyatsky>

Academic adviser:

Dr. Joel Lanir // <https://www.linkedin.com/in/joel-lanir-9a327825>